Structure-based Molecular Design: Identification of Modulators of the NCS1:D2 Interaction

Thesis submitted in accordance with the requirements of the

University of Liverpool

For the degree of

Doctor in Philosophy

by

Victoria Elizabeth Pedder

September 2015

Abstract

The interactions between specific proteins (PPIs) is known to be critical for numerous biological processes, implicating them in many pathological conditions, thus modulation of PPIs has substantial therapeutic potential. The complexity, topography and, in some cases, the hydrophobic nature of the PPIs presents a considerable challenge. One important PPI of therapeutic interest, that has been implicated in the treatment of bi-polar and schizophrenia disorders, occurs between neuronal calcium sensor 1 (NCS1) and the dopamine receptor 2 (D2).

The research detailed in this thesis describes the application of structure-based drug design (SBDD) to select small molecule compounds for synthesis and biophysical assessment against the NCS1 D2 target. The use of a structure-based drug design method, has been seen in previous PPI studies and uses a combination of techniques, including computational modelling used in conjunction with “hit identification” and ”hit to lead” optimisation processes in a drug discovery pipeline. The biophysical analyses of the first generation synthesised were hampered by problems associated with limited aqueous solubility, restricting the determination of accurate affinity values. Thus a second generation of ligands were developed with addition of solubilising groups to the scaffold, based on that of the compound 1-benzyl-N-((2-methoxy-4,6-dimethylpyridin-3-yl)methyl)-3,5-dimethyl-1H-pyrazole-4-carboxamide (**Inhibitor 2**). The second generation of compounds displayed improved aqueous solubility, in particular; 1-(4-chlorobenzyl)-3,5-dimethyl-N-((5-(morpholine-4-carbonyl)pyridin-3-yl)methyl)-1H-pyrazole-4-carboxamide (**Inhibitor 5**), presented the most promising hit.

A fragment based approach was also investigated, adapting the SBDD approach by developing a computational pipeline to select 28 compounds from a library of 1137 for biophysical screening. A two-step biophysical screening protocol was developed; employing high throughput NMR techniques, five fragments were identified alongside a hit fragment candidate 5-methyl-3-phenyl-1H-pyrazole (**4.21**).

This research presents two applications of an *in silico* screening protocol able to identify ligands targeting PPIs. Through verification *via* biophysical techniques, a number of compounds were determined as hits however, no affinity for the target was determined. This project highlights that despite some successes, many challenges remain in the development of targeting PPIs with small molecules.

Table of Contents

Acknowledgements

There have been many people throughout the course of my PhD that I would like to thank, firstly my supervisors Dr Andrew Carnell, Professor Lu-Yun Lian and Dr Neil Berry. Your guidance, support, encouragement and patience with me has been invaluable to my experience over the past four years.

I have also been very privileged to work in two excellent labs within the Department of Chemistry and the Institute of Integrative Biology; Firstly to the members of Lab 1.84 passed and present, to Katie and Lee, thank you for your constant ridicule over our well scheduled tea breaks, I believe they call it character building?! Lab C and the NMR centre, there's far too many of you to name, thank you for welcoming me to biology, for being so understanding and putting up with this clumsy chemist.

To the costa crew; Dr Hannah Davies, Dr Jill Madine and Dr Marie Phelan without our weekly meetings over tasty beverages I do not know how I would have coped with the scientific and moral dilemmas of the past four years. Marie has been especially helpful not just whilst working within the NMR centre, but even when I was working in chemistry. For having the patience of a saint with my continuous questioning and for reading the entirety of this thesis I will be forever grateful.

The lady who inspired me to study Chemistry, who never gave up on me during my time at school and sixth form Mrs Shan, you were always so kind, understanding and believed in my abilities when perhaps others were not so sure. I was lucky enough to be one of the many pupils who over the course of your career you took under your wing and helped achieve their goals. Thank you for all that you have done for me and many more, there is not enough Guinness in the whole of Ireland that could repay your kindness to us all.

Finally, perhaps the most important thankyou is to my family especially my mum and dad, you have always had faith in me even when I have not had faith in myself. It can't be easy having a PhD student for a daughter, especially one that lives 222 miles away who likes to talk to you about her day in work every day and still gets lifts home for the holidays!! I can honestly say that without your love and support through university I would not have been able to get this far and because of that I dedicate this thesis to you both.

List of Abbreviations and Definitions

Å	Ångström
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor
APS	Ammonium persulfate
ASP	Astex statistical potential
Bcl-xL	B-cell lymphoma-extra large
Bak BH3	Pro apoptotic peptide with a BH3 domain
D2R	Dopamine receptor 2
d	Doublet
dd	Doublet of doublets
DCC	*N,N*-dicyclohexylcarbodiimide
DCM	Dichloromethane
DIPEA	*N, N*-Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DPPF	1, 1'-Bis (diphenylphosphino) ferrocene
EDCI	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EF Hand	Helix loop helix motif
ELISA	Enzyme linked immunosorbent assay
ES	Electrospray
Fmoc-Cl	Fluorenylmethyloxycarbonyl chloride
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
frq1	Frequenin
FtsZa	Filamenting temperature-sensitive mutant Z A
GRK2	G-Protein coupled receptor kinase 2

HATU	1-Bis(dimethylamino)methylene-1H,1,2,3-triazolo(4,5b)pyridinium3-oxihexafluorophosphate
HDM2	Human protein double minute
HOBt	Hydroxybenzotriazole
HPV	Human papilloma virus
HTS	High throughput screening
HRMS	High resolution mass spectroscopy
HSQC	Heteronuclear single quantum coherence
IC_{50}	Half maximal inhibitory concentration
IL-2	Interleukin-2
IR	Infrared spectroscopy
ITC	Isothermal titration calorimetry
K_D	Disassociation constant
m	Multiplet
MDM2	Mouse double protein minute
MeCN	Acetonitrile
MeI	Iodomethane
MeOD	Deuterated methanol
MeOH	Methanol
m/v	Mass per volume
mw	Molecular weight
N	Normal (concentration units)
NCS	Neuronal calcium sensor
NCS1	Neuronal calcium sensor 1
NEt_3	Triethylamine
NMR	Nuclear magnetic resonance

OMe	Methoxy
PDB	Protein data bank
$Pd_2[dba]_3$	Tris(tribenzylideneacetone)dipalladium(0)
$Pd[dba]_2$	Bis(dibenzylideneacetone)palladium(0)
PGDS	Prostaglandin D synthase
pik1	Phosphatidylinositol 4-kinase
PLP	Piecewise linear potential
PPI	Protein-protein interactions
ppm	Parts per million
P53	Tumour suppressor protein
q	Quartet
RMSD	Root mean squared deviation
s	Singlet
SAR	Structure activity relationship
SBDD	Structure based drug design
SDF	Standard delay format
SDS-PAGE	Sodium dodecyl sulfate- polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
TEV	Tobacco etch virus
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMED	Tetramethylethylenediamine
TNF	Tumour necrosis factor
VDW	Van der Waals
TOCSY	Total correlation spectroscopy
v/v	Volume per volume

WT	Wild type
w/v	Weight per volume
ZipA	Z interacting protein A
α	Alpha
β	Beta
1-D	One dimensional
2-D	Two dimensional
3-D	Three dimensional

Chapter 1
Introduction

1 Introduction

1.1 Neuronal Calcium Sensor Proteins

This thesis is concerned with a class of Ca^{2+} binding proteins that has been implicated in a number of neurological disorders. Such proteins can be subdivided into two classes (a) buffers and (b) sensors, depending on their affinity for Ca^{2+}, binding off rates and functionality.[1] Members of the Ca^{2+}buffer class include parvalbumin, calbindin and calretinin however, they shall not be discussed further as it is members of the sensors that are of interest to this research.[2]

The neuronal Ca^{2+} sensors (NCS) are a family of EF-hand containing Ca^{2+} binding proteins.[1] The NCS proteins can be subdivided into five groups (A - E), depending upon their appearance in evolution, or their sequence similarity and have been identified in numerous organisms from yeast to man.[1,3] The members of the five sub-classes include: class A (NCS proteins), class B (neurocalcin and hippocalcin), class C (recoverin), class D (guanylyl cyclase activating proteins, GCAPs) and finally class E (potassium channel inactivating proteins, KChIPs).

Members of the NCS family are able to bind Ca^{2+} with varying affinity and as such have different purposes, NCS proteins regulate signal transduction in the brain and retina.[4] Recoverin is only expressed in the retina, where it has been found to be a sensor in rod and cone cells whereby it controls the desensitisation of rhodopsin.[4] Many other NCS proteins are expressed in the brain and have diverse functions for example, neurocalcins and visinin-like proteins are implicated in synaptic plasticity through their regulation of guanylate cyclase and nicotinamide acetylcholine receptors.[5]

Other proteins such as KChIP, hippocalcin and NCS1 are involved in the control of neuronal excitability through their interactions with ion channels.[6–8] Within the NCS proteins, there is a large degree of sequence identity which would suggest that they would have conserved three dimensional structures.[3,4] This is observed in both the X-ray crystallography or NMR structures of Ca^{2+} bound un-myristoylated NCS proteins (Figure 1.1).[4]

The evolutionary conservation of the NCS proteins can be observed, all are approximately 200 amino acid residues in length and possess four EF hands (four helix loop helix motif's).[3,9] Of those four, only three are able to bind Ca^{2+} (although recoverin and KChIP1 can only bind two).[3,9] Furthermore it is the first EF hand in the *N*-terminal region of all NCS proteins that is inactive in the binding of Ca^{2+} and of

the 14 mammalian NCS proteins 11 are *N*-terminally myristoylated which enables them to anchor to cell membranes, albeit using different mechanisms.[3,10]

Despite the similarities, the ability of the different NCS proteins to recognise different sets of physiological ligands suggests that they must have some distinguishing structural difference, which therefore makes the NCS family potential druggable candidates.[4]

No PPI inhibitors have as yet been identified for NCS1 a member of this important class of proteins.

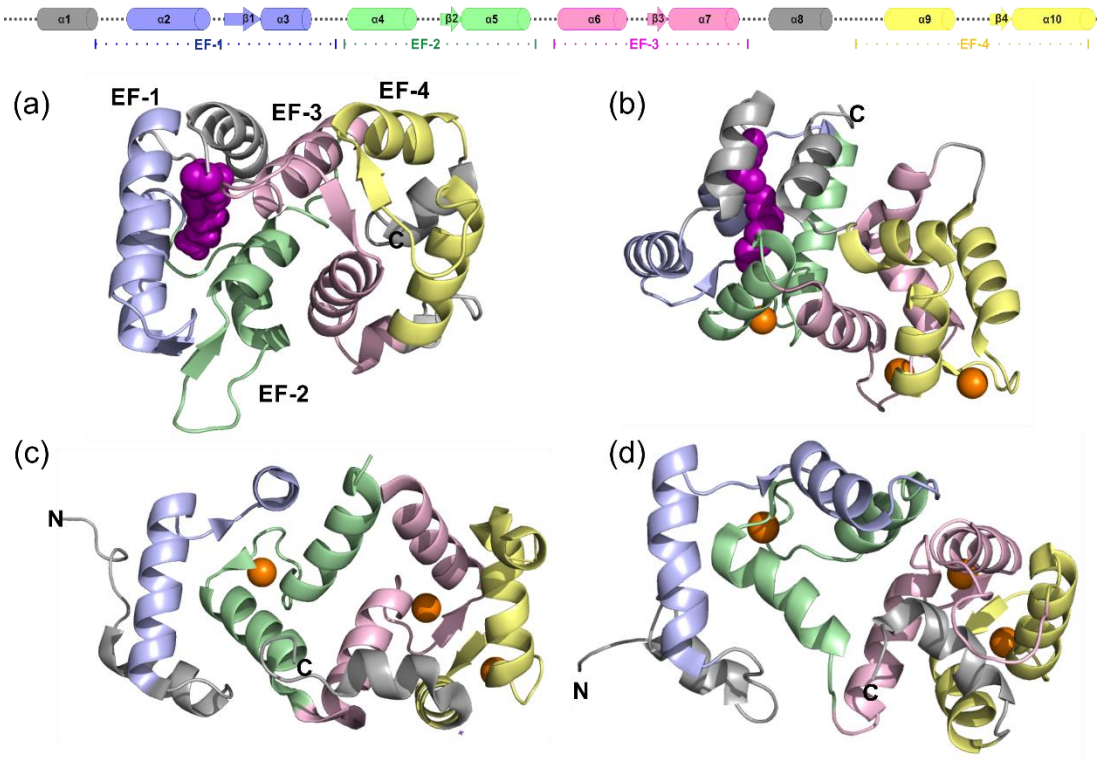


Figure 1.1- A comparison of some of the different NCS protein structures- (Top) a generalised representation of the secondary structure of a member of the NCS proteins colour in accordance to the four EF hands, light blue (EF-1), light green (EF-2), light pink (EF-3), and pale yellow (EF-4). **(a)** Myristoylated mammalian recoverin (PDB 1iKU [11]) the myristoyl group coloured in purple, the four EF hands can be seen in light blue (EF-1), light green (EF-2), light pink (EF-3), and pale yellow (EF-4) **(b)** Myristoylated mammalian GCAP1 (PDB 2r2l [12]) the myristoyl group coloured in purple, with the Ca^{2+} as orange spheres, again the four EF hands can be seen in light blue, light green, light pink, and pale yellow. **(c)** Rat NCS1 (PDB 5AEQ [13]) the four EF hands are coloured in accordance with the previous structures and the Ca^{2+} as orange spheres. **(d)** *S.cerevisae* frq1 *(*PDB 1FPW [14]) the four EF hands are coloured in accordance with the previous structures and the Ca^{2+} as orange spheres.

1.1.1 NCS1 and the dopamine receptor D2

Mammalian NCS1 also known as human frequenin (Figure 1.1 c) has been implicated in numerous physiological functions including the regulation of neuro transmitter release, synaptic plasticity, learning and memory function.[15–18] An investigation into the interacting partners, indicate that NCS1 is involved in many protein-protein interactions in both a Ca^{2+} dependent and independent manner.[19] Amongst the

interacting proteins identified are phosphatidylinositol-4-kinase (PI4K) and its yeast orthologue pik1, the dopamine receptors D2R and D3R, ARF1, the interleukin receptor accessory protein like-1 IL1RAPL1, TRPC5 channels and InsP(3) receptors.[19–27] Of these wide ranging interacting partners, one of considerable interest and importance is the dopamine receptor D2 due to its role within the central nervous system.

The dopamine receptor D2R is a member of the G-protein coupled receptors (GPCRs) of which there are five (D1-D5) that can be subdivided into two families; the D1 like family which include D1 and D5 and the D2 like family including D2, D3 and D4.[28] Comparisons between the dopamine receptor D3 and D2 lead to the prediction that D2 is composed of seven transmembrane α helices, which are connected through three intracellular and three extracellular loops.[29] Alongside an amphipathic α helical *C*-terminal domain as seen with D3, the amino acid sequence in the *C*-terminus of D3 differs from that of D2 by only one residue (Serine D3 for Histidine D2, Figure 1.2).[29]

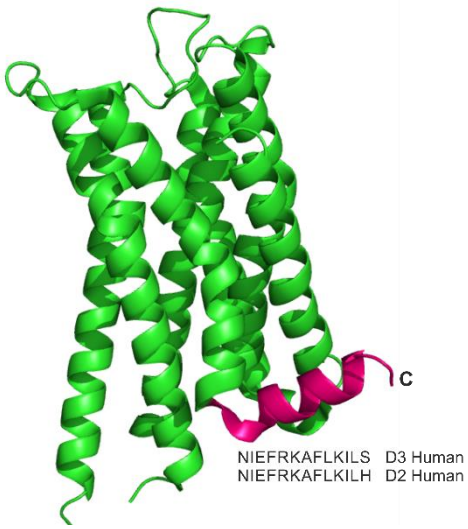


Figure 1.2- The structure of the dopamine D2/D3 receptor (PDB 3PBL) with the C-terminal helix coloured magenta and the amino acid sequence annotated. The D2R peptide is predicted as being α helical, the C-terminal region of D3 which only differs from the sequence of D2R peptide by one residue is also known to form an amphipathic α helix as shown here.[27,29]

The biological implications of the dopamine receptor D2 include dopaminergic signalling within the central nervous system, addictive behavioural disorders, bi-polar disorder and schizophrenia.[30,31] Interestingly, the expression levels of NCS1, are increased in patients with bi-polar disorder and schizophrenia when under an antipsychotic regimen such as haloperidol **1.1** or clozapine **1.2** (Figure 1.3).[30]

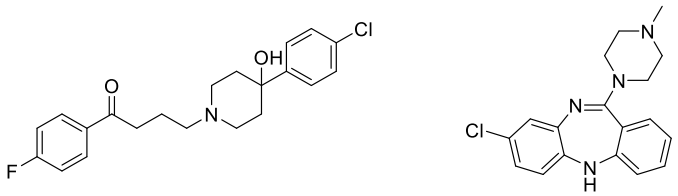


Figure 1.3- The chemical structures of the antipsychotic drugs haloperidol **1.1** and clozapine **1.2**.

The mechanism for the interaction of NCS1 with D2 is both direct and indirect in nature; the Ca^{2+} sensitive interaction between NCS1 and D2 causes the desensitisation of the D2 receptor towards dopamine to be reduced (the direct effect Figure 1.4).[32] However, the interaction between NCS1 and the glycoprotein coupled receptor kinase 2 (GRK2) blocks the GRK2 mediated phosphorylation of the D2 receptor (this is the indirect effect Figure 1.4).[32] The Ca^{2+}dependent nature of the NCS1 interactions are regulated through changes in the Ca^{2+} concentration, this occurs through the release of Ca^{2+} from intracellular stores and Ca^{2+} channels (Figure 1.4).[32]

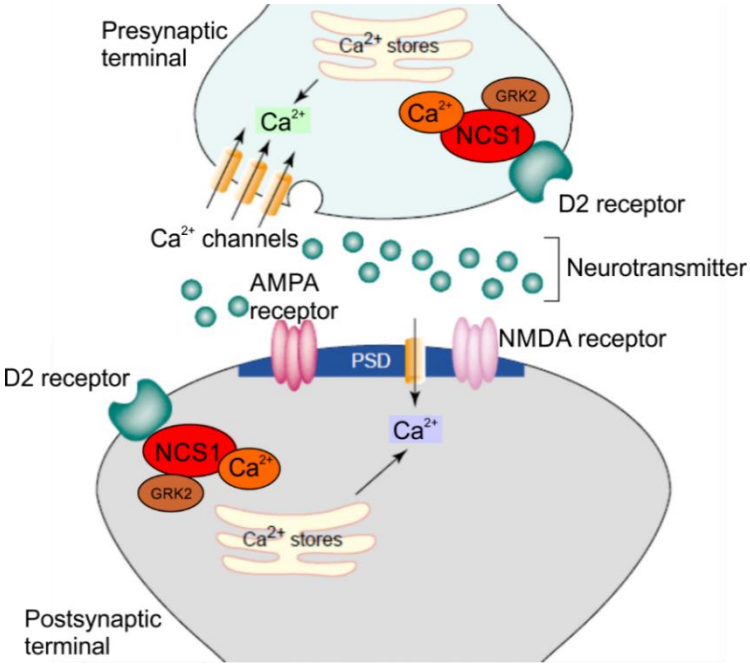


Figure 1.4- A representative view of the interaction between NCS1 and dopamine receptor D2 (D2R) at the synapses (adapted from [32]). The Ca^{2+} sensitive interaction between NCS1 and D2R causes attenuation of the D2 receptor desensitisation. The interaction of NCS1 with the glycoprotein coupled receptor kinase 2 (GRK2) is also Ca^{2+} dependent and it blocks the GRK2 mediated phosphorylation of the D2 receptor. The Ca^{2+} dependent nature of NCS1 is regulated through changes in the Ca^{2+} concentration through the release from intracellular stores and Ca^{2+} channels.

The physiological relevance of the NCS1-D2 interaction has heralded it as a promising target for molecular intervention. No PPI inhibitors have as yet been identified for NCS1.

With the structures of numerous members of the NCS proteins already determined using X-ray crystallography or NMR spectroscopy, it was suggested that the NCS1-D2R interaction could occur in a similar manner.[33] It was observed in the previous cases, that the interaction of a ligand or peptide occurred in a large hydrophobic cleft that was formed by a number a hydrophobic residues.[4,34–37] The residues in question were known to be conserved throughout the NCS family, suggesting that this mode of binding could be adopted across the NCS proteins.[38] However, the specificity of each interaction would occur through differences in the size of the cleft most likely in relation to the size of the binding partner.[38] In the case of NCS1 it was found that the presence of Ca^{2+} caused a large hydrophobic cleft to become exposed as with previous NCS proteins, this furthered the assumption that this could be the site for the interaction of a protein or ligand.[33]

More recently an extensive study into the different crystal structures of NCS1 complexes and truncated Ca^{2+} bound proteins by Lian *et al.* elucidated a number of interesting and significant features (Figure 1.5).[13] Utilising a number of techniques including X-ray crystallography, it was confirmed that two D2R peptide molecules bound to NCS1, as had been postulated by previous computational and NMR spectroscopic based studies.[13,38]

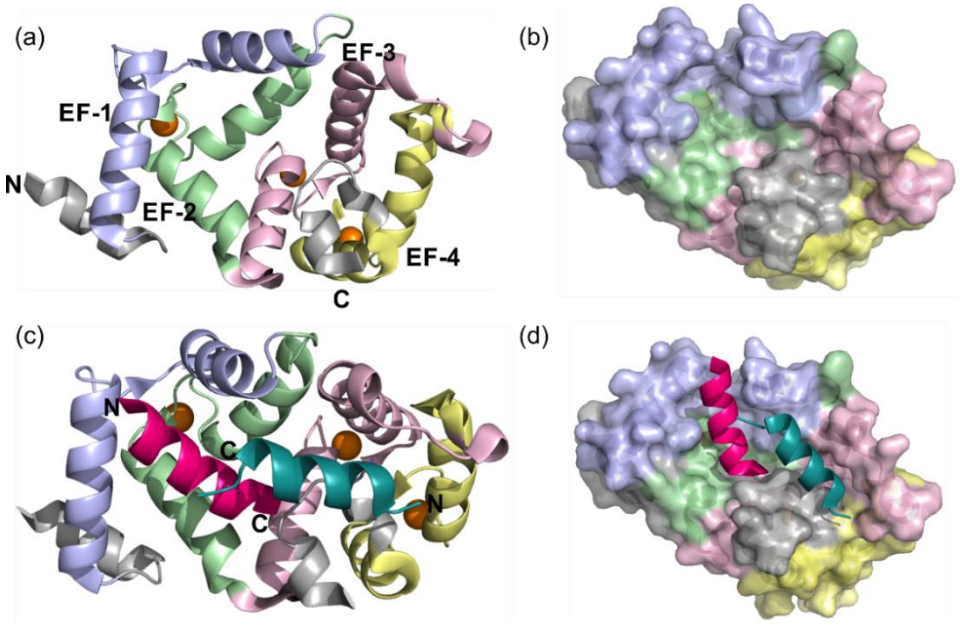


Figure 1.5- Structure of rat Ca^{2+} NCS1 complexed with dopamine receptor D2R peptide (PDB 5AER). (a) Cartoon representation of the NCS1 structure when in complex with the dopamine receptor D2R peptide omitting the D2 peptide. The EF hands are coloured accordingly EF-1 pale blue, EF-2 pale green, EF-3 pale pink and EF-4 cream the three Ca^{2+} orange spheres. **(b)** The molecular surface of NCS1 coloured as before, with the hydrophobic ligand binding site observed running down the centre. **(c)** Cartoon representation of the NCS1 D2R peptide complex, NCS1 coloured as before with the two C-terminal helices of two D2R peptides observed in magenta and teal, Ca^{2+} orange spheres, the two peptides bind to NCS1 independently of one another. **(d)** Molecular surface representation of the NCS1 D2R peptide complex coloured as before (adapted from [13]).

A number of key hydrophobic residues within the conserved hydrophobic groove were implicated in the binding interaction between the protein and the peptide (Figure 1.6).[13] The binding of NCS1 to the two peptides can be split into two specific regions of the protein, the *N*-lobe and *C*-lobe, with one peptide molecule binding to each region separately. The two D2R peptides were found to bind to NCS1 with their *C*-termini orientated towards the centre of NCS1 (Figure 1.5). Those implicit residues of NCS1 found to undergo an interaction within the *N*-lobe include Trp30, Phe34, Phe48, Ile51, Tyr52, Phe55, Phe85 and Leu89.[13] Interestingly when compared to other NCS protein and peptide complexes, the evolutionary conservation of these residues indicate their importance for binding within the complexes.[13] In comparison, the *C*-lobe was found to make fewer intermolecular contacts which include Phe54, Ile128, Tyr129, Met131, Val132, Val136, Leu138, Ala182, Tyr186 (Figure 1.6 b and Figure 1.7). However, despite having fewer contacts to the D2R peptide, it appeared to undergo significant structural changes upon binding. These changes include the induced-fit structural stabilisation of one side of the hydrophobic crevice between residues Val132 - Leu138. In the uncomplexed protein (Figure 1.1 c) this region was found to be unstructured, however in the presence of the D2R peptide (Figure 1.5 b) a helical conformation is adopted.[13] Interestingly it is this *C*-lobe made up of the EF-3 and EF-4 regions that are the least conserved within the NCS family members. It could be suggested that the residues within this region that are stabilised upon D2R peptide binding could give rise to target specificity and as such could be key for therapeutic purposes.

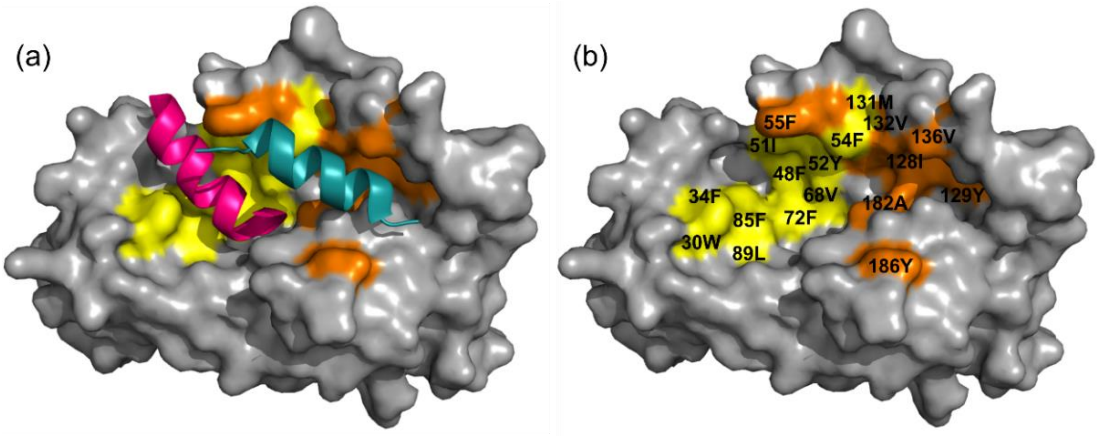


Figure 1.6- Structure of rat Ca^{2+} NCS1 in complex with D2R peptide-the key hydrophobic residues (PDB 5AER). (a) Molecular surface of NCS1 coloured grey, the key hydrophobic residues that interact with the C-terminal helix of the D2R peptides 1 (magenta) and 2 (cyan) coloured yellow and orange respectively. **(b)** As with (a) but the two D2R peptide helices removed and the key residues annotated (Adapted form [13]).

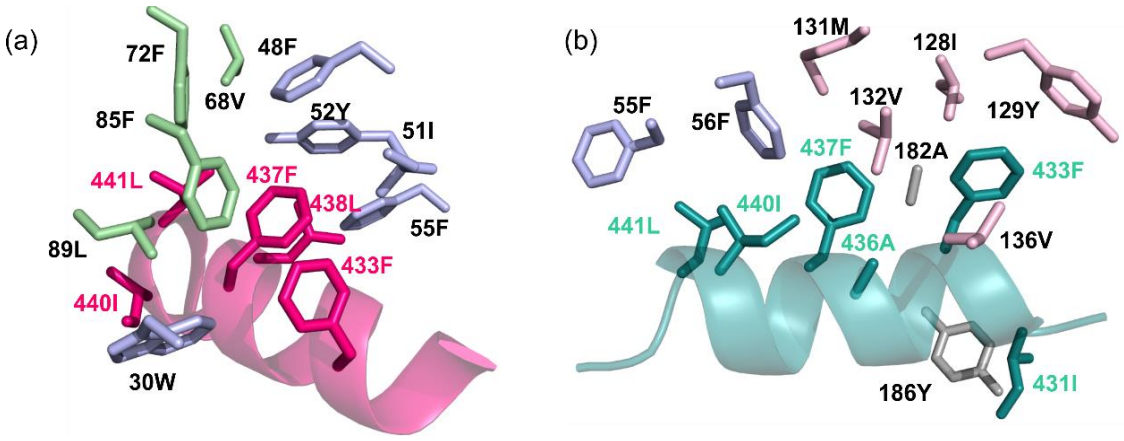


Figure 1.7- The Intermolecular interactions of rat Ca^{2+} NCS1 with the D2R peptide. (a) A detailed zoom of the hydrophobic interactions of the sidechain residues from the N-lobe of NCS1 (coloured as Figure 1.5) with the C-terminal helix of D2R peptide 1 (magenta). **(b)** A detailed magnification of the hydrophobic interactions of the sidechain residues from the C-lobe of NCS1 (coloured as before) with the C-terminal helix of D2R peptide 2 (teal).

1.1.2 Frq1 and pik1

As well as mammalian NCS1, it has been found that the yeast *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* genomes encode a protein that shares a sequence homology of greater than 60% with mammalian NCS1.[4] These homologs have different roles in their respective genomes. For example in the fission yeast *S. pombe* the homologue known as Ncs1 is responsible for the regulation of sporulation and Ca^{2+} tolerance.[39,40] In the budding yeast *S. cerevisiae,* the orthologue is known as frq1 and it is essential for cellular growth.[26] A key protein-protein interaction has been found to occur between frq1 and phosphatidylinositol 4-kinase (pik1) (Figure 1.8). The pik1 protein is fundamental for vesicular trafficking in the late secretory pathway, nuclear functions and cytokinesis, frq1 is necessary for optimal pik1 activity.[34,41]

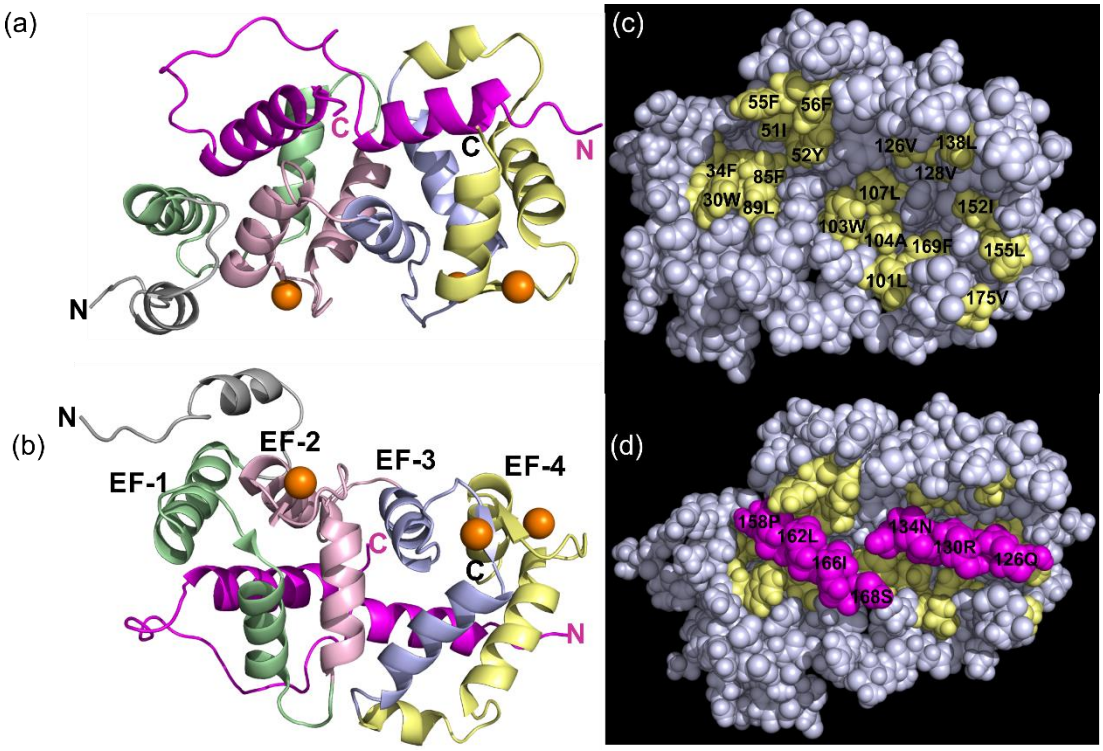


Figure 1.8- **The frq1 pik1 complex**. **(a)** View from the binding interface and **(b)** the binding face rotated 180°, both cartoon representations of NMR structure of the interaction between yeast frq1 and pik1. The four EF hands of frq1 can be seen in light green, light pink, light blue and cream, with the two alpha helices of pik1 in magenta. **(c) and (d)** Space fill representations from the binding interface **(c)** of frq1 alone with its key hydrophobic residues in yellow **(d)** frq1 and pik1 with the key residues of pik1 annotated (adapted from [34]).

The pik1 peptide is composed of two α-helices, the *N*-terminal helix from residues 127 - 135 and the *C*-terminal helix 156 - 167, the two helices are connected by a twenty residue loop.[34] The stoichiometry of binding deduced from NMR studies was defined as 1:1, in contrast to that of the 2:1 stoichiometry determined for the D2R peptide NCS1 interaction .[34] As with the NCS1 D2R peptide interaction the protein-protein interaction between frq1 and pik1 involves two lobe regions of frq1 interacting independently with two α helices of pik1. The hydrophobic residues of the *N*-terminal helix of pik1 interacts with the *C*-terminal residues of frq1.[34] Similarly, the hydrophobic residues on the *C*-terminal helix of pik1 contact the *N*-terminal hydrophobic groove of frq1 (Figure 1.8).[34]

It was discovered through the results of NMR studies by Strahl *et al.* that in the complex of frq1 and pik1 the terminal residues of frq1 at both the *N* and *C* terminus are solvent exposed and disordered. This lead to the possibility that the *N*-terminal myristoyl group of uncomplexed frq1 may actually be involved in recruiting pik1 to the membranes.[34]

Work by Lim *et al.* found that frq1 underwent a large conformational change induced through Ca^{2+} binding, whereby the myristoyl group was extruded, this caused the

hydrophobic residues in the *C*-terminal to become solvent-exposed enabling it to interact with pik1.[4] Furthermore through their NMR analysis, the interaction between the *C*-terminal helix of pik1 and frq1 was found to be more stable in comparison to the corresponding interaction of the *N*-helix of pik1.[34] A number of hydrophobic residues on the *C*-terminal helix of pik1 were found to make extensive contacts with the aromatic and other hydrophobic residues on frq1 (Figure 1.9 a).

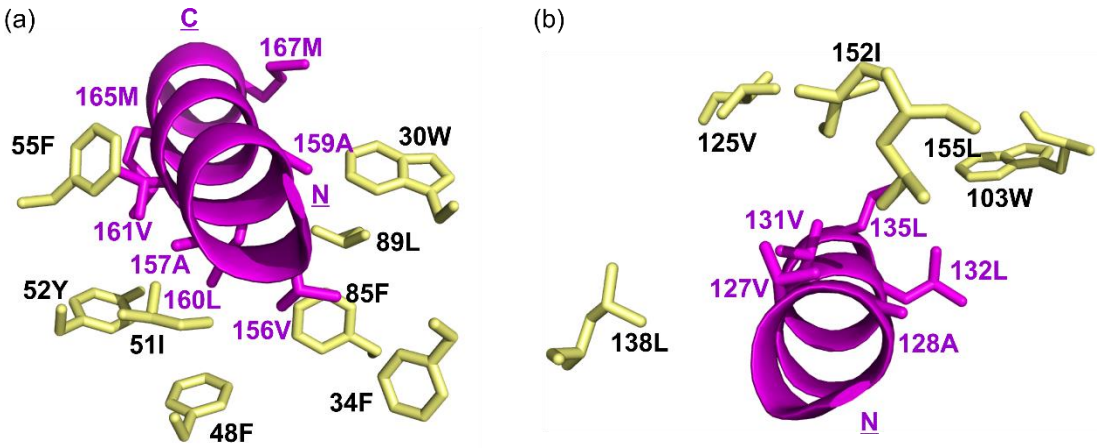


Figure 1.9- **(a)** Close up view of the *C*-terminal helix of pik1 (magenta) with the sidechains of the *N*-terminal hydrophobic groove of frq1 (pale yellow). **(b)** Close up view of the *N*-terminal helix of pik1 (magenta) with the sidechains of the *C*-terminal hydrophobic groove of frq1 (pale yellow) (adapted from [34]).

The network of strong and intricate intermolecular hydrophobic interactions explain the high affinity of the interaction between the two proteins.[34] It suggested that the interaction of the *C*-terminus of pik1 was the anchoring point between the peptide and the protein, facilitating further interaction between the second N-helix of pik1 and frq1.[34]

Frequenin has been found to have a high degree of evolutionary conservation with a similar secondary structure to the mammalian NCS1, with the EF hand motifs EF-2, EF-3 and EF-4 bound to Ca^{2+} ions.[41] Another similarity drawn between the two homologues and other NCS proteins is that they have a hydrophobic pocket, found at the surface between two of the EF-hand-containing lobes as well as the solvent exposed N and C termini.[41]

The structural similarities and the fact that the two orthologues NCS1 and frq1, both interact with pik1, suggest that NCS1 may be responsible for the regulation of the activity of pik1 in mammalian cells.[34] The knowledge that the hydrophobic residues within the *C*-terminal helix of pik1 make considerable contact with aromatic and other hydrophobic sidechains of frq1, could be useful when designing inhibitors.[34] Such important hydrophobic residues include Val156, Ala157, Ala159, Leu160, Val161 and

Met165.[34] This small cluster of hotspot residues could be used to generate a starting template or three dimensional representation, from which a computational structure-based drug design (sbdd) method could be applied. The sbdd method would allow for the selection of novel compounds, able to replicate the area of 3-D space that the hotspot residues represent for a pharmacophore based search. The compounds would subsequently be synthesised and tested using a variety of biophysical methods.[34]

There are a variety of different computational applications used in structure-based drug design process (Section1.4), and a number of protein-protein interactions (PPIs) have been targeted therapeutically, using computational applications in development of drug-like compounds.

1.2 Protein-Protein Interactions (PPIs) as Drug Targets

PPIs are involved in nearly all biological processes and cellular function is therefore reliant on their regulation.[42] Due to their ubiquitous nature and central role, there has been an increasing interest in targeting the interface between two interacting proteins as a potential therapeutic target.[42,43] Inhibition of protein-protein interactions has seen significant recent advances, although it does present a number of challenges because of the nature of protein-protein interaction surface.[43] The surfaces involved in PPIs are typically large, often flat with complicated topography, and a large portion of the surface area of the interfaces are buried.[44,45] Despite these large interfaces there may only be a small number of influential residues essential for a high affinity binding interaction [46–48], thereby, reducing the actual challenges posed when developing small molecule inhibitors and making the targeting of protein-protein interactions a realistic and viable drug development strategy.

A typical passively absorbed drug profile follows Lipinski's 'Rule-of-five' which includes a MW < 500 Da. Due to the nature of protein-protein interfaces molecules which are targeted at them tend to have a high molecular weights.[43,49] They also tend to have an increased hydrophobicity, a higher number of hydrogen bond donors, hydrogen bond acceptors and a higher ring complexity in comparison to other drugs.[50]

At present excluding peptidomimetics there is no well-established structure-based approach for the design of PPI modulators using defined computational pipelines. There is also a general lack of small molecule starting points for drug design as there are very few known naturally occurring small molecules that bind to protein

interfaces.[44,50] It can be difficult to distinguish between true binding from artefactual binding and current small molecule compound libraries tend to be biased towards certain protein classes, making them more specific to certain targets and less general.[44,50]

Despite the difficulties PPI targets present, there have been a number of notable success stories with several drugs are already in clinical use and some others are being evaluated in clinical trials.[51]

1.3 The inhibition of PPIs past and present

Research into the design and synthesis of small molecule inhibitors has been conducted for a number of key protein-protein interactions that are believed to play a vital role in many biological processes. Due to the central role of PPIs the number of systems studied is increasing very rapidly, with some of the most interesting being, studies on binding to IL-2, HPVE2, ZipA, TNf, HDM2 and Bcl-X_L.[43] Here we shall briefly discuss some of the approaches developed to target these interactions, in particular those of HDM2 and Bcl-X_L (Sections 1.3.5 and 1.3.6).

1.3.1 IL-2

IL-2 (Interleukin-2) is a cytokine that is of particular interest due to its role in the activation of killer T-cells in an immune response and has been linked to the rejection of tissue grafts.[52] The design of a series of small molecules able to bind to the receptor of IL-2, was conducted by Sunesis Pharmaceuticals, without prior knowledge of the structure of the IL-2 bound to its receptor (IL-2R).[53] The initial hit to lead process started with a small peptidomimetic designed to mimic the structure of IL-2 binding region; however it was found to bind to IL-2 itself rather than the IL-2 receptor and so was not taken any further.[54]

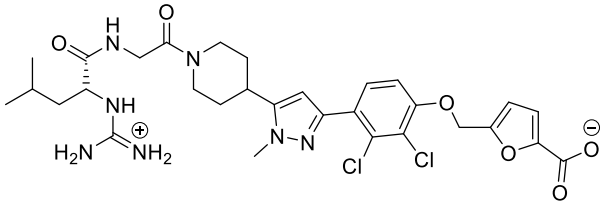
Interestingly, the subsequent discovery process was carried out utilising a combination of techniques and using the original unsuccessful compound as a template starting point. The compound was divided into its four constituent fragments, from which the subsequent structure activity relationship studies (SAR) elucidated one key fragment. Optimisation of this fragment was carried out through a number of steps including further compound library screening processes to evolve the target compounds.[43,54]

To ensure a greater success rate at each separate stage the lead compounds were characterised using a number of assays including enzyme-linked immunosorbent assay's (ELISA); this confirmed binding of the small molecule to the IL-2R hotspot.[54]

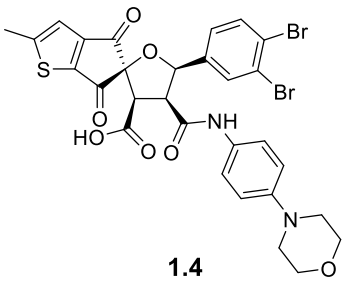
The most successful small molecule designed was SP4206 (**1.3** Figure 1.10) which interestingly, does not resemble the IL-2R.[53] In fact it is believed that if the crystal structure of the IL-2 receptor complex had been used in the design process then SP4206 would not have been discovered, highlighting the occasional unpredictability and sometimes serendipitous nature of drug discovery.[43]

1.3.2 Interactions involving the human papilloma virus HPV

The protein-protein interaction between the viral transcription factor E2 and the viral helicase E1 is important in the life cycle of the human papilloma virus (HPV).[55] HPV is associated with cervical cancer and warts, and small molecule treatment regimens are available with limited success.[56] Therefore the development of such drugs to target HPVE2 and prevent its interaction with E1 is a key target for pharmaceutical companies.[56] A research group at Boehringer Ingelheim was able to utilise High Throughput Screening (HTS) to identify a class of compounds known as indadiones.[56] These compounds were found to disrupt moderately the protein-protein interaction between E2 and E1; however, further target optimisation and medicinal chemistry efforts led to the development of the **1.4** (Figure 1.10). This compound exhibited a lowered IC_{50} value of 6 nM, with a higher ligand efficiency than the original binding partner target E1.[55,57,58] Thought to be due to the ability of the compound to bury its hydrophobic surface area, rather than spanning across the vast protein-protein interface.[55,58]



1.3



1.4

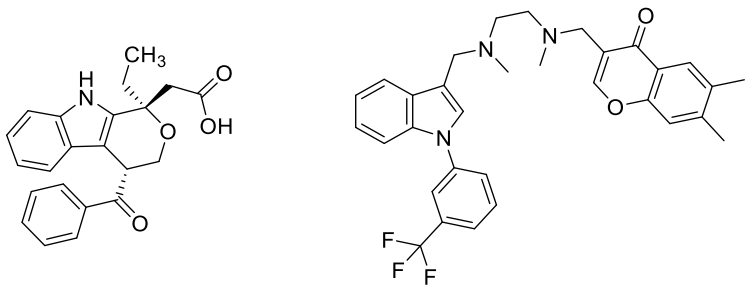
Figure 1.10- Structures of two developed small molecule inhibitors (**1.3**) SP4206 an inhibitor developed for targeting IL-2 (**1.3**) the so called "Compound 23" developed by Boehringer Ingelheim for the HPV: E2 interaction (structures adapted from [53,55,58])

1.3.3 ZipA and FtsZa

In another study the interaction between ZipA and FtsZa was examined, ZipA is a membrane anchored protein that interacts with a homologue of eukaryotic tubulin, FtsZa through terminal carboxy domains.[59] These proteins were found to be the main constituents of a structure known as a septal ring found in gram negative bacteria, and is responsible for their separation and replication.[59] The interaction between the proteins was studied using X-ray crystallography, NMR spectroscopy was employed to screen a diverse set of 825 compounds.[60] The NMR study found seven hits that bound to ZipA at the same site as FtsZa and is an example of early drug design processes.[60] However, despite this early indication that ZipA may be "druggable" further medicinal chemistry and SAR were unable to generate any small molecules able to penetrate deep within the surface of ZipA which is in stark contrast to the findings of studies on IL-2, HPVE2, HDM2 and Bcl-X_L.[43,60,61] This is an excellent example of the challenges that are involved in targeting complex interactions associated with two interacting proteins.

1.3.4 TNF

Tumour Necrosis Factor (TNF) is also a cytokine and is an important biological target due to its involvement in the inflammatory response.[62] It is the interaction between TNF and its receptors, TNFR1 and TNFR2, that the small molecule inhibitors are designed to disrupt, these targeted therapeutics have been approved for the treatment of arthritis.[62] Fragment screening lead to the development of a group of small molecules, that were found to interfere with TNF binding, through the displacement of one of the three monomers that constitute the TNF-α homotrimer.[63] However due to their moderate affinities, these compounds are not considered as potential drug candidates.[63]



1.5 **1.6**

Figure 1.11- The chemical structures of two small molecules, (**1.5**) was developed to target the ZipA interaction and the TNFα inhibitor (**1.6**) (adapted from [60] and [63]).

1.3.5 Case Study 1: Developing small molecule modulators of P53 and MDM2 over the decades…

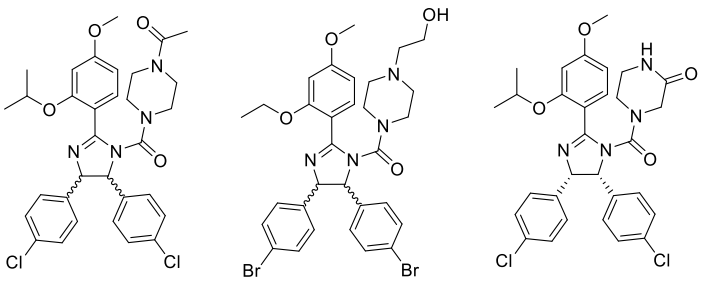
An extensively investigated protein-protein interaction is that between the tumour suppressor gene p53 and HDM2.

HDM2, also known as the human protein double minute, was found to be an excellent target in the treatment of cancer.[64] The initial research into MDM2, the mouse orthologue, found that it bound to the tumour suppressor protein p53, increasing its degradation and blocking its transcriptional activity which results in tumour suppression.[65] The interaction of MDM2 with p53 also results in the nuclear movement of p53 into the cytoplasm and away from its site of action.[66]

The design and synthesis of compounds to modulate this interaction has been of substantial interest over a number of decades. However despite keen research interest, the development of small molecule modulators was not successful for a long period of time, indicative of the complexity involved in the design of modulators for PPI targets.[67]

The interaction between the two proteins MDM2 and p53 was found to involve of a small segment of the *N*-terminus transcription activation domain of p53 (residues 13-19) and a hydrophobic pocket within MDM2.[68] Furthermore X-ray crystallography, indicated the primary involvement of three sidechains of the hydrophobic residues; Phe19, Trp23 and Leu26, which form a short α helix on p53 located within a deep pocket in MDM2 (Figure 1.13).[67,69]

One of the earliest success stories of inhibiting the two interacting proteins was reported by Vassilev *et al.* who developed a series of *cis*-imidazoline compounds, more commonly known as Nutlins (Figure 1.12).[70] High-throughput NMR spectroscopy screening of a diverse library of compounds, followed by extensive optimisation strategies from the initial leads resulted in three racemic hit compounds Nutlin-1 and Nutlin-2. However interestingly, of the two enantiomers of Nutlin-3 only one is active and this is known as Nutlin-3a.[70] The three Nutlin compounds were found to disrupt the p53-MDM2 interaction with IC_{50} values of 260 nM, 140 nM and 90 nM, respectively.[70] Determination of the crystal structure of the complex between MDM2 and Nutlin-2, indicated that the compound was able to mimic the sidechains of Phe19, Trp23 and Leu26 on the p53 helix, through investigations of the corresponding interaction sites on MDM2. Since its initial publication Nutlin-3a has been evaluated for its therapeutic potential for cancer treatment.[70]



1.7 Nutlin-1 **1.8** Nutlin-2 **1.9** Nutlin-3a

Figure 1.12 The chemical structure of Roche's *cis*-imidazoline compounds the Nutlins, designed to inhibit the p53 MDM2 interaction. The two racemic compounds Nutlin-1 and Nutlin-2 and the active enantiomer Nutlin-3a.

Cummings *et al.* at Johnson and Johnson used the proposed three key residues from the interacting sequence of a p53 9mer segment (Figure 1.13), as a template in the design of their benzodiazepine derived inhibitors.[67] When screened against their biological target, these benzodiazepine derivatives presented half maximal concentration (IC_{50}) values ranging from 125 μM - 0.4 μM.[67] A crystal structure of one of the benzodiazepine inhibitors with MDM2 (PDB 1T4F), indicated it was able to successfully mimic the three sidechain residues, this structural data further supported a competition assay with the native peptide.[67,71]

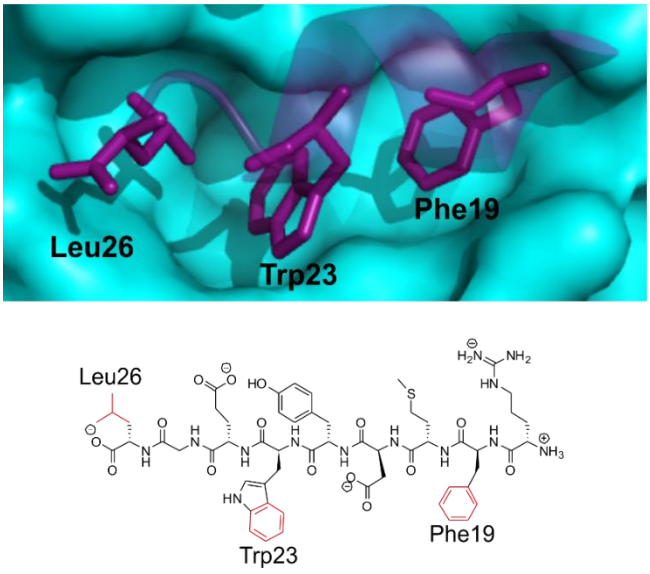
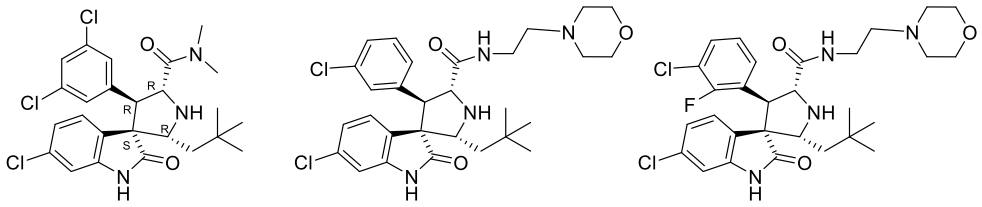


Figure 1.13- The three hydrophobic sidechain residues of the p53 helix (purple) highlighted in the deep pocket of MDM2 (cyan) (PDB 1T4F), against the corresponding 9mer segment of p53 with the sidechains highlighted below (adapted from [67]).

Ding *et al.* elucidated a key hydrogen bonding interaction between the NH of the indole ring of sidechain from Trp23 to the MDM2 backbone using x-ray crystallography.[68] Therefore the design strategy of small molecule inhibitors developed by Ding *et al.* involved searching for chemical moieties that could mimic this interaction.[68] They found that in addition to the indole ring an oxindole could

mimic the hydrogen bonding and hydrophobic contributions of the Trp23 sidechain and this was therefore used as a core structure for the starting point of the structure-based strategy.[68] From here a substructure search was conducted to identify natural products that contained the oxindole ring, among those identified were spirotryprostatin which contained a spiro-oxindole core.[68]

Molecular modelling studies using computational docking, found these compounds fitted poorly into the MDM2 cleft due to steric clashes, but the spiropyrrolidone ring presented a rigid scaffold from which the residues of Phe19 and Leu26 could be mimicked through the addition of hydrophobic sidechains.[68] Candidate compounds were then designed through the variation of the sidechains, their most potent inhibitor **1.10** (Figure 1.14) was found to have a K_D value of 86 nM with MDM2.[68] It was an effective inhibitor of cellular growth in cancer cells with WT p53 and exhibited selectivity over cancer cells with deleted p53, with minimal toxicity to healthy cells and presented a promising lead compound for optimisation.[68]



1.10 **1.11** **1.12**

Figure 1.14- The chemical structures of three non peptidic inhibitors of p53 and MDM2, a product of the structure-based design and synthesis process approach developed by Ding *et al.* (adapted from [68,69]).

The three residue mimic was also used to design other inhibitors; however through the use of fluorescence polarisation assays it was discovered that despite the high binding affinities of these molecules they were significantly less potent.[69] This indicated the presence of other interactions, confirmed with the discovery of a fourth residue Leu22 involved in the binding interface (Figure 1.15).[69] Molecular modelling techniques such as computational docking using the crystal structure was then used to optimise **1.10** bound to MDM2.[69] The findings suggested that a chain extension at the carbonyl end could interact with the desired fourth residue.[69]

The computational docking programme GOLD was the next step in the process, predicting binding modes of several different chain extensions. The fact that the Leu22 residue was not buried within a hydrophobic pocket, but rather solvent exposed meant that the group could experiment with different polar moieties that could also aid the solubility profile of the compound.

Through fluorescence polarisation (FP) assays, the initial target **1.11** (Figure 1.14) displayed an inhibition constant (K_i) of 13 nM and an increased potency, six times greater than **1.10**.[69] The synthetic route to obtain **1.11** utilised an asymmetric 1, 3-dipolar cycloaddition reaction as the main step in the stereospecific synthesis and is shown in Scheme 1.1.

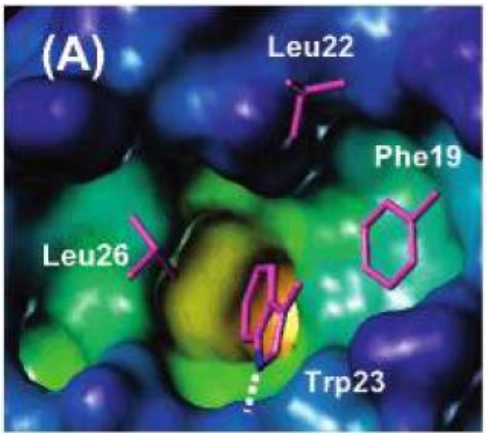
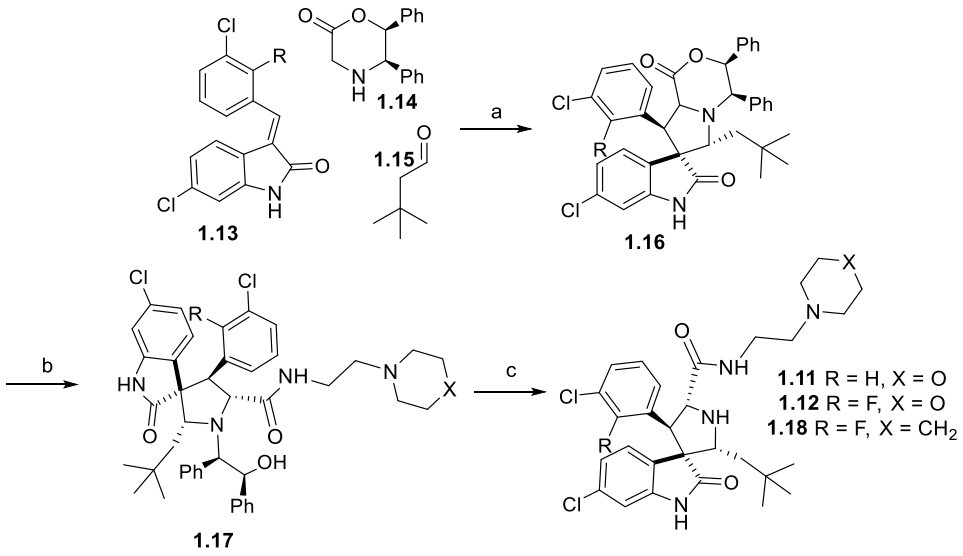


Figure 1.15- X-ray structure of p53-MDM2 complex (PDB 1YCR). In addition to the three residues Phe19, Trp23 and Leu26, Leu22 was also found to play a role in the interaction. MDM2 cavity is coloured according to its depth, the deep buried region coloured in yellow with the external solvent exposed regions in blue and the sidechains of the key residues in magenta (adapted from [69]).

Further medicinal chemistry driven optimisation of **1.11**, involved attempts to improve the metabolic stability of the compounds through the incorporation of fluorine groups on the benzene lead to a reduction in their potency. However, development of another ligand **1.12** from the scaffold of **1.11** with the incorporation of a fluorine group, resulted in a K_i of 3 nM. When Ding *et al.* used their fluorescence polarisation assay to compare their newly designed inhibitor to that of Hoffman la Roche's most successful candidate **1.9** Nutlins 3a, they found that **1.12** was 12 times more potent.



Scheme 1.1- The three step stereospecific synthetic route to spiro-oxindole derived p53 MDM2 modulators 1.11 and 1.12 developed by Ding *et al.* (adapted from [69]). **(a)** 4 Å molecular sieves, toluene; **(b)** 2-morpholin-4-yl-ethylamine/THF or 1(2-aminoethyl)-piperadine/THF, RT; **(c)** $Pb(OAc)_4$, $CHCl_2$-MeOH (1:1), 0 ˚C.

This is an excellent example by Ding *et al.* of the use of rational drug design and guided by computational modelling used alongside chemical synthesis and structural biology to develop and optimise small molecule inhibitors of PPIs.

More recently an extensive investigation into the structure-based rational design behind the discovery of a novel piperidinone series of p53 MDM2 inhibitors was reported by Rew *et al.* (Figure 1.16).[72] Their investigative studies were based on a rigid core that was capable of holding two aromatic rings, to again target the key residues of p53. Initial results led to the development of **1.19** a 1,3,5,6-tetra substituted piperidinone which had an IC_{50} of 2.42 μM.[72] By simply changing the configuration of the acetic acid component and converting the diaryl from *cis* to *trans* followed by isolation of the active enantiomer **1.20,** increased the potency up to 70 fold.

This new optimised inhibitor was found to be capable of forming electrostatic interactions between the acetic acid component and an imidazole sidechain of MDM2.[72] They subsequently reported that optimisation processes of the *N*-alkyl substituent on **1.20** resulted in compounds that were designed to occupy the Phe19 pocket, improved solubility through increasing the polarity of the compound and creating van der Waals contacts to MDM2.[72] Furthermore the discovery of a third hit compound **1.21** came from optimizing the piperidinone scaffold so that the *trans*-diaryl conformation was stabilised with substitution of a methyl moiety on the same carbon as the acetic acid functionalisation (C3 Scheme1.2).[72] The most potent

compound **1.22** was obtained through optimisation of **1.21**, it demonstrated an excellent efficacy with a binding constant (K_D) of 0.4 nM and was selective for its target with an improved oral bioavailability and pharmacological studies predicted it to have a long half-life in humans.[72]

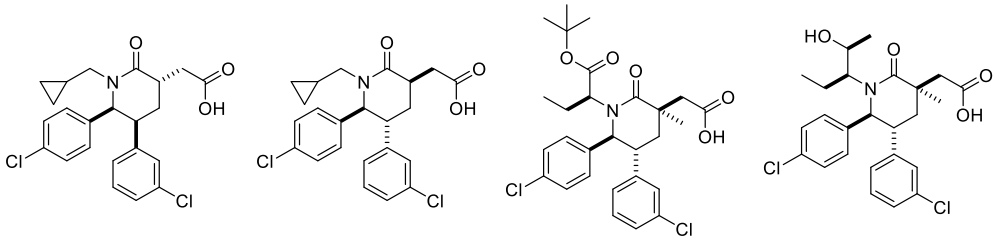
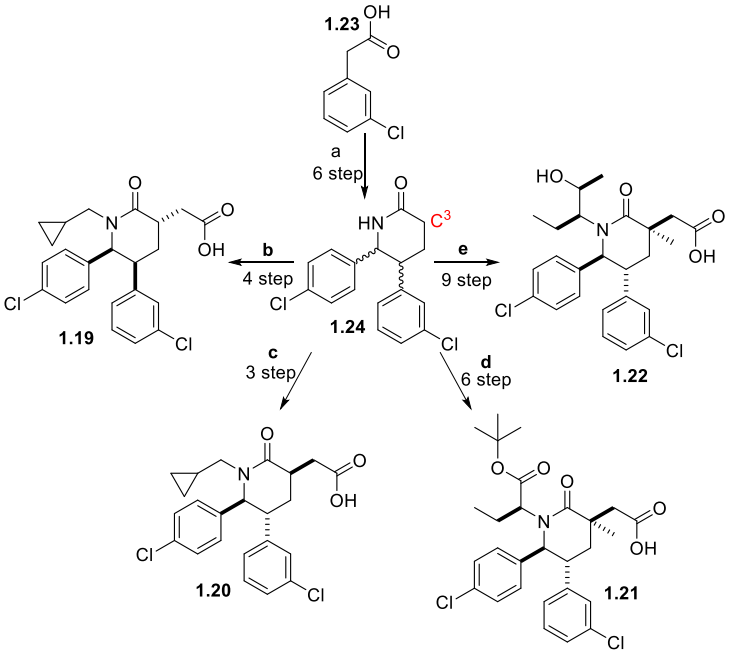


Figure 1.16- The five piperidinone inhibitors developed by Rew *et al.* The development process involved the initial hit compound **1.19**, optimisation processes led to **1.20**, which in turn was optimised to compound **1.21** and ultimately led to the most potent compound **1.22** which had a K_D of 0.4 nM.

The synthetic procedures employed in this vast hit to lead optimisation project followed multi step syntheses which are summarised in Scheme 1.2. From the simple starting compound 2-(3-chlorophenyl) acetic acid **1.23** a six step synthetic procedure (a) involving such processes as conjugate addition, $NaBH_4$ reduction and the Staudinger reaction lead to **1.24** in its *cis*-diaryl form.

A further three steps converted the *cis*-diaryl **1.24** into the first hit compound the *cis*-diaryl 1,3,5,6-tetrasubstituted piperidinone **1.19**. The second hit compound **1.20** was generated from the enantiomer of *trans*-diaryl **1.24**, the reactions include *N*-alkylation, followed by allylation at the C3 position and finally oxidative cleavage of the alkene at C3 using catalytic amounts of ruthenium tetroxide. The six synthetic procedures to obtain **1.21** included some of the following protection of the amide nitrogen of the *trans*-aryl piperidinone **1.24**, methylation at C3, deprotection followed by *N*-alkylation and oxidative cleavage of the alkene as before.

A further three successive reduction and oxidation reactions using the Des-Martin periodiane from an analogue of **1.21**, resulted in the final and most successful lead compound **1.22** in a yield of 75%. This is an excellent example of a scaffold based synthesis that is able to utilise a wide variety of synthetic organic chemistry techniques to achieve optimisation of hit to lead. Although the processes may be lengthy and not high throughput, the synthesis which revolves around further functionalisation of a core scaffold, works in sequential manner that is efficient at producing the hit compounds numerous analogues for biological testing.



Scheme 1.2- The general synthetic approach to obtain the hit compounds 1.19, 1.20, 1.21 and 1.22. The processes involved multistep syntheses originating from the parent starting material 2-(3-chlorophenyl) acetic acid **1.23** which is converted to the piperidinone core **1.24** in a six step synthesis 87 – 88% yield **(a)**. The piperidinone core is presented as a general scaffold without any stereochemistry, however to achieve the desired stereochemistry of **1.19, 1.20**, **1.21** and **1.22** the reagents and conditions of **(a)** were altered as necessary. The *cis*-diaryl **1.20** is converted to the first hit compound the *cis*-diaryl 1,3,5,6-tetra substituted piperidinone **1.19** in a four step synthesis 79% yield **(b)**, the second developed compound the *trans*-diaryl carboxylic acid **1.20** was synthesised from *trans*-diaryl **1.22** in a three step synthetic procedure 89% **(c)**. A six step synthetic procedure 47 – 80% **(d)** was used to transform the *trans*-diaryl scaffold of **1.24** to the carboxylic acid **1.21** and finally a nine step procedure **(e)** was employed to achieve the most successful hit compound **1.22** 80% yield.[72]

Further to these examples of the numerous compounds developed as MDM2 inhibitors, at least two compounds have more recently reached clinical development. MDM2 inhibitors include an oral treatment for advanced stage solid tumours (Johnson and Johnson, Phase 1), and R7112 (Hoffmann-La Roche, Phase 1) an oral treatment for hematologic neoplasms and advanced solid tumours.[73]

1.3.6 Case Study 2: Targeting the interaction of Bcl-X_L and BaK

The second case study to be discussed is that of Bcl-X_L and BaK, within the Bcl2 family, Bcl-X_L inhibits programmed cell death and BaK promotes apoptosis.[74] The interactions between the two proteins controls the sensitivity of the cell towards apoptosis through the antagonism of the two different functions.[74] Bcl-X_L contains two central hydrophobic α-helices which are in turn surrounded by five amphipathic helices, it binds tightly to a 16 amino acid region (BH3) of BaK (residues 72-87) thereby inhibiting the promotion of cell survival (Figure 1.17).[74]

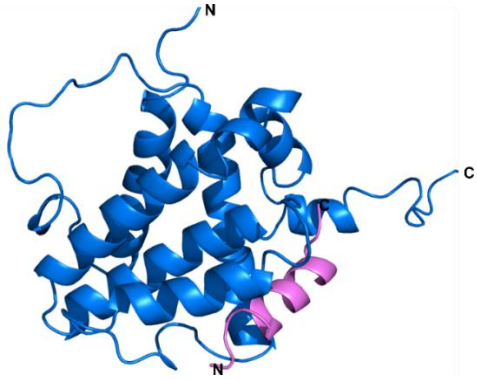


Figure 1.17- The Bcl-X_L (blue) / BaK (Violet) complex derived from NMR studies by Sattler *et al.* (PDB ID 1BXL, adapted from [74]).

Considerable research has been conducted by Hamilton *et al.* into targeting the interface with small molecule antagonists to promote apoptosis in unhealthy cells.[75] Several approaches were used to identify small molecule inhibitors of Bcl-X_L, including oligomeric aryl-aryl and aryl-carboxamide components, in these instances the substituent's would mimic the distance and angular projection of the sidechains from the α helix of its binding partner BaK3.[75] Hamilton utilised intramolecular hydrogen bonds along with the steric interactions to help position the peripheral functionalities and instil a rigidity to the structure to aid pre-organisation at a molecular level.[75] Through the use of chemical synthesis alongside structural biological techniques such as, X-ray crystallography, fluorescence polarisation (FP), isothermal titration calorimetry (ITC) and NMR spectroscopy studies, Hamilton was able to develop enaminone **1.25**, benzoylurea **1.26**, terphenyl **1.27** and terepthalamide **1.28** scaffolds.[75–77] These four novel scaffolds were able to successfully mimic the residues of an α-helix (Figure 1.18).[75–77]

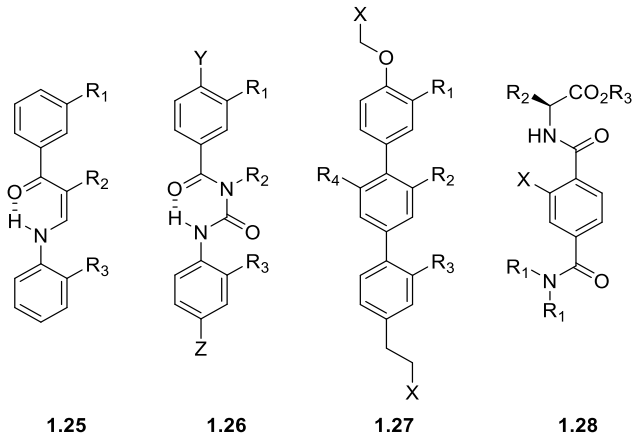
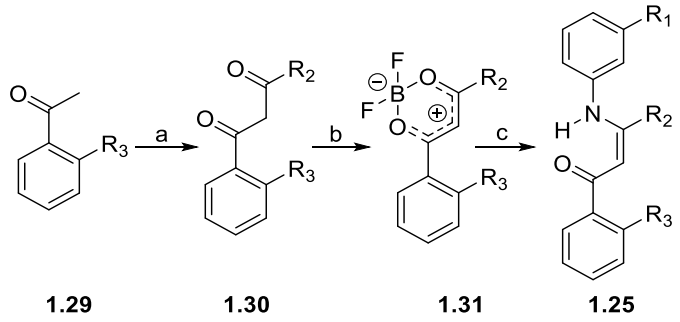


Figure 1.18- The α-helix small molecule mimetics developed by Hamilton and co-workers. enaminone **1.25**, benzoylurea **1.26**, terphenyl **1.27** and terepthalamide **1.28** scaffolds where R1, R2 and R3 are the different positions of functionalisation.[75–77]

The BaK3 helix has been shown through crystallographic evidence to contain four key hydrophobic residues Val74, Leu78, Ile81 and Ile85 whose sidechains play a vital

role in its interaction with Bcl-X_L.[77] The original terphenyl derivatives **1.27** were the first compounds synthesised to mimic projection of these sidechains, they adopted a staggered conformation with substituents that projected their groups at the i, i+4 and i+7 positions as with the α helix of BaK3.[76] The synthesis was developed using *O*-alkylated phenols and sequential Suzuki aryl-aryl cross coupling reactions.[76] The synthesis was lengthy and the compounds were found to be extremely hydrophobic with a predicted *logP* = 9.25 and low aqueous solubility *logS* = -10.76.[77]

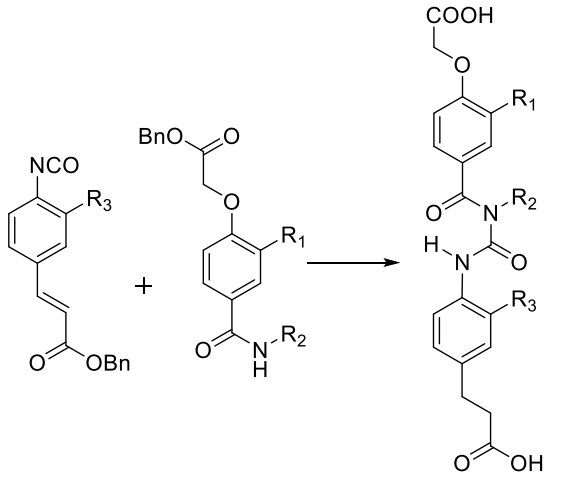
The terephthalamides **1.28** were developed as a second generation of small molecules and were simplified in terms of structure; they were also designed to have a better solubility profile. [77] The two terphenyl rings of **1.27** were replaced by two functionalised carboxamides, which were able to replicate the planar nature of the phenyl rings through the rigidity and restricted rotation about the amide bond.[77] The synthesis of these compounds was simplified when compared to that of the terphenyls, involving *O*-alkylation and amide couplings and was found to be much more efficient.[77] In comparison to the terphenyls, terephthalamides predicted *logP* = 4.42 and a solubility constant logS= -5.35 were much more desirable.[77] The enamine scaffolds **1.25** were synthesised using an even faster and efficient synthesis (Scheme 1.3). This involved initial Claisen condensation reaction to give the di-ketone **1.30**.[78] Subsequent reaction with BF_3OEt_2 resulted in the 1,3-diketonatoboron difluoride complex which in the final reaction was reacted with *m*-toluidine to produce the desired enamine **1.25**.[78]



1.29 **1.30** **1.31** **1.25**

Scheme 1.3- The synthetic approach to the enamine scaffold 1.25. The synthetic approach involves a Claisen condensation reaction of **1.29** using **(a)** NaH, EtOAc, EtOH and ether to give the diketone **1.30. (b)** The subsequent reaction of the diketone with BF_3OEt_2 to give the di-fluoride compound **1.31** in 84% yield, **(c)** which is finally reacted with m-toluidine to give the target compound the enamine **1.25** in 70% yield**.**

Finally, the benzoylurea compounds **1.26** were designed based on the enamine scaffold, the synthetic approach involved the synthesis of two fragments; an amide and an isocyanate which were coupled to form **1.26** (Scheme 1.4).[75] The two substituent fragments were synthesised using similar synthetic approaches involving Wittig reactions, reductions, iodonations, *O*-alkylations and Heck reactions.[75]



1.32 **1.33** **1.26**

Scheme 1.4- Coupling of an amide and isocyanate substituents to give the benzoylurea scaffold **1.26** in 43% yield (adapted from [75]).

These examples of the extensive investigations conducted by Hamilton *et al.* not only produced efficient small molecule inhibitors of the Bcl-X_L protein. They were also able to provide a detailed understanding of the small molecule bound Bcl-X_L structure through informed computational docking programmes and numerous biophysical techniques.[75–77]

With these case studies of different PPIs and the alternate approaches employed to target them, it is clear that the process involves the combination of many techniques and iterative of design, synthesis and testing.

The successful compounds were developed using computational modelling to predict aspects such as binding poses and solubility. A variety of medicinal chemistry techniques were used alongside diverse synthetic organic chemistry, to realise a range of scaffolds that were able to either mimic the positions of amino acid sidechains, or target certain hotspot residues. When used in conjunction with structural knowledge of the biological target and a number of biophysical techniques, it lead to improving the initial hit compounds towards lead molecules and in some cases, ultimately to the clinic.

This interdisciplinary approach was successful in targeting protein-protein interactions implicated in numerous biologically relevant viruses, diseases and disorders such as cancer, arthritis, gram negative bacterial infections and HPV.

1.4 Computational applications of structure-based drug design

In designing inhibitors for protein-protein interactions, the complex nature of the systems involved means that conventional high throughput screening methods may not be an effective way to generate hit compounds. This may be observed in a high number of false positives that occur due to the hydrophobic nature of many binding sights. Indeed it is clear that a new pre-screening strategy is necessary.[79]

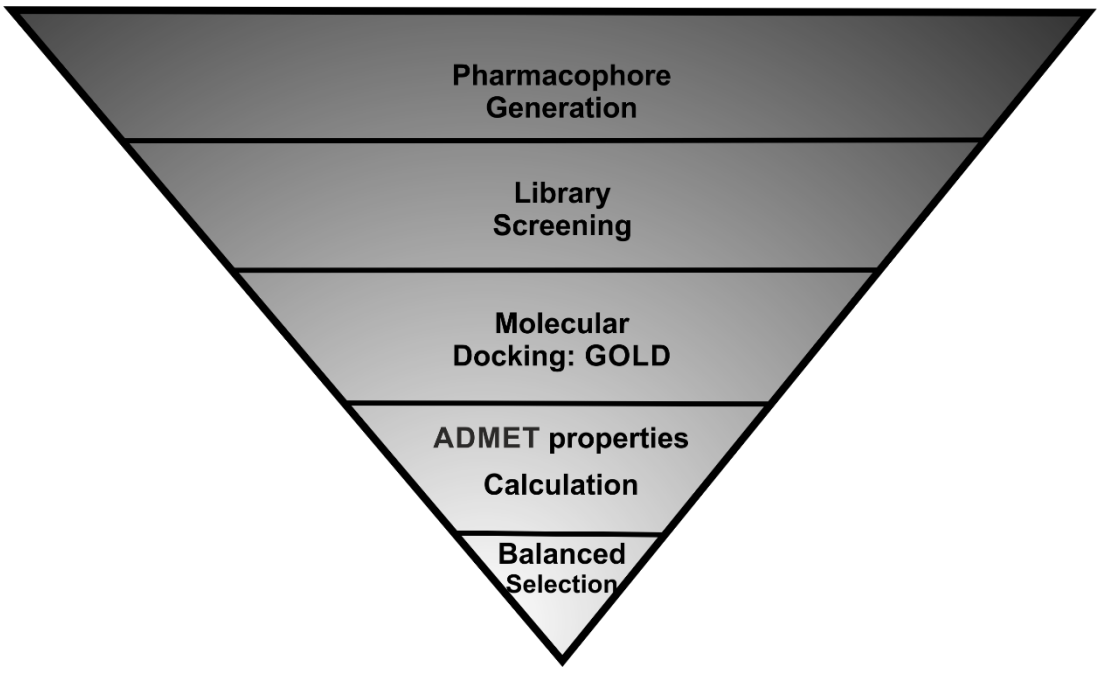


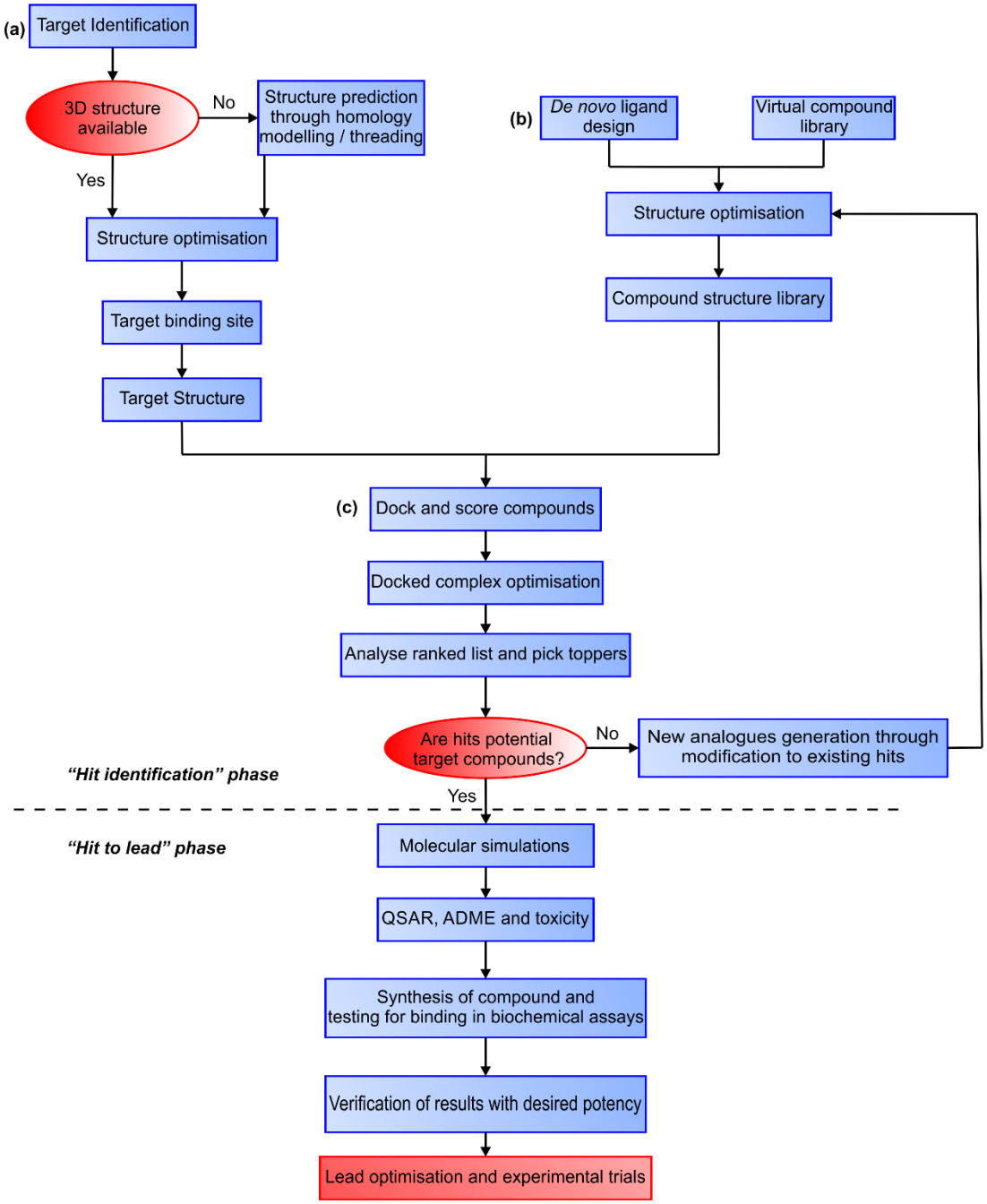
Figure 1.19- The computational workflow applied in this thesis, following a general structure-based drug design process. Pharmacophore generation is used to screen a library of compounds to replicate an area of three dimensional space, the compounds are then docked into the target and their ADMET properties subsequently calculated. The information is then combined and a balanced selection protocol is undertaken to select the best possible candidate for synthesis. At each stage the number of potential hits is reduced.

As discussed previously, computational applications have found use in the development of drug-like compounds to target protein-protein interactions. As seen with the development of inhibitors to target the Bcl-X_L and BaK3 interaction (Section 1.3.6), Hamilton *et al.* designed four novel scaffolds which are able to mimic the amino acid sidechains at positions i, i+4 and i+7 of the BaK3 helix. Once two scaffold structures had been designed, the computational programme QikProp was used to calculate certain drug-like parameters, such as the solubility profile of each ligand to compare and contrast the two and determine if the scaffolds were suitable.[77] The conformations of the compounds in solution were then investigated using the molecular modelling facility MacroModel, wherein they employed energy minimisation with the MM2 force field.[77] Subsequently the poses or orientations generated from

MacroModel were then superimposed onto the native BaK3 α-helix, to compare the positions of the functionalised substituents, to the i, i+4 and i+7 amino acid sidechain positions on the BaK3 α-helix.[77]

Ding *et al.* also used the computational docking programme Gold, to calculate the binding pose of their initial hit spiro-oxindole **1.10** (1.3.5) and from that develop a number of analogues including their lead compound **1.12**.[69] These simple computational processes aided in the selection of the compounds to synthesise and is an example of how computational techniques can aid success when used in the process of targeting PPIs.

The use of a structure-based drug design method, as seen in previous PPI studies (1.2) uses a combination of techniques. The methodology involves computational modelling, used in conjunction with “hit identification” and ”hit to lead” optimisation processes in a drug discovery pipeline (Scheme 1.5).[80,81] Furthermore, to improve this computational screening process, it is more favourable to start with the best selection of ligands to be computationally analysed and selected for synthesis. Application of a pharmacophore filter could be one such way to do this.



Scheme 1.5- The structure-based drug design pipeline. The flow process applied in the SBDD protocol follows two phases, the "hit identification" phase and the "hit to lead" phase. A number of processes occur in the hit identification phase and these include **(a)** Target identification and preparation of the binding-site. **(b)** Preparation of the compound library used to screen hits for the target. **(c)** Molecular docking methods, analysis of the compounds predicted binding interactions with the target. This is followed by the hit to lead phase, selection of a lead compound (adapted from [82]).

1.4.1 Pharmacophore Modelling

The term pharmacophore was defined by Ehrlich as 'a molecular framework that carries the essential features responsible for a drug's biological activity.[83] The composition of the pharmacophore encompasses the steric and electronic characteristics necessary to ensure optimal interactions between a molecule and target ligand.[83] Selection of the correct three-dimensional disposition of chemical

features is important for the development of a successful pharmacophore model.[84] There are two types of pharmacophore model, ligand-based and structure-based. The first works in the absence of a 3-D structure of the macromolecular target; it takes a group of active molecules and compares their common 3-D features.[83] A structure-based pharmacophore model works directly with the 3-D structure of a macromolecular target or macromolecule-ligand complex.[83]

1.4.1.1 Ligand-based pharmacophore modelling

The ligand-based method was not used for the purposes of this thesis and so shall not be discussed in great detail. However, briefly the method involves two main steps. Firstly the conformational flexibility of each ligand has to be represented by creating conformers, and secondly the ligands then have to be aligned and the common features extracted to generate the pharmacophore.[83] This approach does not come without a number of challenges such as the ability to model the ligands flexibility and selecting the best method to align the molecules.[85,86] Perhaps a more challenging question is finding a way to select the best group of compounds to generate the pharmacophore, ensuring that the selection is diverse and not biased.[86]

1.4.1.2 Structure-based pharmacophore modelling

In the work presented here structure-based modelling was used. This involves analysis of the complementary chemical features of the active site and the spatial and chemical relationships, allowing a pharmacophore model with selected features to be built.[83] It can be further subdivided into two categories: (i) macromolecule-ligand-based whereby the complex structure is available to determine the hotspot residues between the two binding partners.[83], and (ii) macromolecule- based which is simply the macromolecule target without the natural ligand.

There are a number of modelling programmes which are able to determine the key residues, either in the complex or within the macromolecule itself and these can be used to generate a pharmacophore.[83] As with the ligand-based model, this approach does have a number of limitations, such as the need for the structure of a macromolecule target or macromolecule-ligand complex. Even if the complex or structures are available there are often a vast number of chemical features that have to be rationalised before a pharmacophore can be generated computationally.[83] To overcome these problems a number of algorithms have been developed with the ability to generate hotspot pharmacophores from the structure of the macromolecule itself. It does so by identifying cavities within the structure and comparing these cavities to the ligand binding pockets of other known complexes.[87]

The structure-based method has proven useful in drug design as the model can be used to generate a pharmacophore from structural information of the target. This leads to a more informed screening process for complimentary ligands that are able to interact at the desired locations on the macromolecule. Generation of the pharmacophore subsequently enables the searching of databases for potential ligands in a process known as 'pharmacophore-based virtual screening'.[83,88]

1.4.2 Pharmacophore-based virtual screening

Pharmacophore-based virtual screening has been applied to a number of disease targets and has proved useful in the hit identification and lead optimisation strategies in drug design.[85,89] It is a process which allows for millions of compounds to be screened *in silico* to find those which fit the specific points defined in three dimensional space. As discussed previously the generation of a pharmacophore can be carried out with the knowledge of a target's structure and as such a number of functional groups including hydrogen bond donors and acceptors, charged moieties and aromatic groups can be included.[90,91]

The process of pharmacophore-based virtual screening has seen a dramatic increase in use for a number of studies on various disease targets; a review by Kim *et al.* highlights recent applications.[88] Examples are given of the success stories of hit generation through pharmacophore-based virtual screening of a variety of targets within numerous disease states such as cancer, viral infections including HIV, diabetes, obesity and Alzheimer's disease.[88] It presents excellent examples of how pharmacophore-based virtual screening can be used and demonstrates how the technique has developed.

Pharmacophore-based virtual screening is not without its problems. It can be time consuming when screening compound libraries that contain a large number of flexible molecules. However, this problem can be overcome by the application of a series of filters, the first of which reduces the number of unsuitable ligands so that the subsequent selection processes are conducted on a more refined selection of compounds.[91]

There are a number of databases / compound libraries that can be used, which are able to utilise a filtering process. ZINCPharmer is the online interface to one such database ZINC; this online resource conducts pharmacophore searches of a library of around 35 million purchasable compounds.[92] The compounds within the library are of explicit conformations of each compound within ZINC, the screening process matches the conformers to the three-dimensional pharmacophore generating hits.

To do this ZINCPharmer uses the open-source Pharmer software which uses an index approach to search more than 176 million conformations of 35 million purchasable compounds, in a very short space of time.[93] The Pharmer method is relatively recent approach to pharmacophore modelling; previous methods such as fingerprint-based or alignment-based were much more time consuming in their searches.[85,94] The interface allows for different filters to be applied to a pharmacophore query, meaning that the compounds can be filtered using e.g. the number of conformations of a single compound and also their molecular properties.[92] Such property filters include molecular weight (mw), root mean squared deviation (RMSD with respect to the desired pharmacophoric features) and the number of rotatable bonds, all of which have been implicated in the design process of drug like molecules.[95] Filtering using RMSD diminishes the number of hits to those who best fit the geometric requirements of the pharmacophore query.[92] Once a set of compounds has been identified, it is possible to download the individual files for the subsequent analysis and processing.

1.4.3 Molecular Docking

The resultant steps in the computational pipeline of a structure-based drug design process, involves molecular docking of the hit compounds from the pharmacophore virtual screening studies (Scheme 1.5). Docking is a prediction of a ligand's conformation and orientation within the target macromolecular structure. There are a number of algorithms that are available within docking programmes to rationalise the interactions between a ligand and a protein.[96] The current available techniques include; molecular dynamics, genetic algorithms, Monte Carlo, point complementarity, distance geometry and fragment based methods, "tabu" and systematic searches.[96] The details of these different algorithms shall not be discussed further here as they are summarised in some excellent reviews.[96–108]

Independent of the variety of algorithms used, there are a number of characteristics any docking programme should have, including the ability to accurately represent the experimental binding modes of a ligand *in silico*.[99] The docking process should be quick and efficient and have the ability to score and rank compounds.[99] GOLD (genetic optimisation for ligand docking) developed by Jones *et al.*, is able to fulfil these prerequisites and for the purposes of this thesis the compounds were docked and scored using a genetic algorithm approach.[99,109]

Nissink *et al.* were able to demonstrate GOLD's ability to replicate the experimental binding modes of 305 ligand protein complexes *in silico* with a 68% success rate.[110]

GOLD is composed of three main constituents: (i) a scoring function that is able to rank different binding modes (Goldscore), (ii) a mechanism by which it places a ligand in the binding site and (iii) the genetic search algorithm that investigates the different binding modes of a ligand.[99] It works by performing automated docking within the volume around the active site of a partially flexible protein, allowing for full acyclic ligand flexibility and partial cyclic ligand flexibility.[99,109]

The Goldscore scoring function is a molecular mechanics based function that is the sum of four explicit terms: the protein ligand hydrogen bond score, the protein ligand Van der Waals score, the intramolecular hydrogen bonds of the ligand and the intramolecular strain of the ligand.[97,99,109] The mechanism by which the ligand is docked into a binding site involves fitting points which are added to hydrogen bonding sites on both the ligand and the protein.[97,99,109] The respective donor and acceptor points are mapped onto the complex and the CH groups of the ligand are matched to the hydrophobic points within the proteins surface.[99] This process is exclusive to GOLD. The genetic algorithm used to search the different possible binding modes of the complex, is programmed to modify a number of specific variables.[99,109] The modifications include: the dihedrals of the protein and the ligand, OH and NH_3^+ groups within the protein and those bonds of the ligand that are rotatable.[99,109] The geometries of the ring systems within the ligand and the position of the fitting points between the ligand and the proteins binding site.[99]

As well as Goldscore the programme enables a variety of different scoring functions to be applied, such scoring functions include Chemscore, Chem Piecewise Linear Potential (PLP) and Astex Statistical Potential ASP.[82] ASP uses a protein-ligand database to measure the atom-atom potential, Chemscore estimates the change in energy that occurs upon ligand binding and Chem PLP is an empirical fitness function which models the steric complementarity between the protein and the ligand.[99,111,112]

1.4.4 Physicochemical parameters

The assessment of a molecule to be “fit for purpose” is to determine the physiochemical parameters which will be used alongside the docking results to select the best possible candidates for synthesis.

The Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) properties of molecules are in a sense generic, meaning they represent ‘the action of the body on the drug’ rather than a more specific process of ‘the action of the drug on the body’.[113] The calculation of these ADMET parameters are essential in the selection

process, optimisation of ADMET properties from lead generation to optimising *in vivo* profiles, is now a process that is carried out prior to candidate selection.[113] Within the ADMET setting, oral bioavailability is an important consideration for the development of a biological molecule as a therapeutic agent.[95] Correlation between the physical properties of an orally administered drug and drug candidate, with successful drug development was conducted by Lipinski and lead to the "rule of five" (RO5). This guideline proposed that an increased chance of drug absorption and permeability, is more likely when the molecular weight is < 500 Da, *logP is* < 5, the number of hydrogen bond acceptors is < 10 and the number of hydrogen bond donors is < 5.[113] It was found that failing a single "rule of five" was acceptable, as there was no statistical significance between not failing any and failing one. However, failing two was detrimental to the oral bioavailability of the drug with 90% of oral drugs in phase II clinical trials meeting these criteria.[113] However, the "rule of five" perhaps should be thought of as a guideline for oral bioavailability, rather than a strict rule for an "ideal drug". In addition to the parameters set by Lipinski, molecular flexibility for membrane permeation was postulated as being desirable by Navia.[114] With the complexation of water to amide bonds as a negative factor for oral bioavailability deduced by Hirschmann.[115] Other negative factors towards oral bioavailability include the number of rotatable bonds within a compound which should be no more than 10, high polar surface area on intestinal absorption, failure to achieve membrane permeation, first pass metabolic processes and active transport from the blood into the excretory system such as to the gut or the liver.[95,116]

There has been great debate on the topic of "ideal drug-like compounds" since the RO5 was suggested by Lipinski *et al.* almost two decades ago in 1997. The guidelines for a good orally bioavailable drug remain clear, however the ideals of "the perfect drug-like molecule" are constantly changing as we learn more about the molecular targets and the ADMET properties of the compounds used to treat them. It has been suggested that perhaps we are too obsessed with the oral bioavailability of a compound and that we should consider drug space from the toxicological view point.[117] A number of guidelines have been developed from this point of view that relate toxicity to other physiochemical properties such as lipophilicity which is directly related to a drugs solubility profile and can be predicted through the computational calculation of the partition coefficient C*logP*.

One such rule has been developed by researchers at the pharmaceutical company Glaxo Smith Kline (GSK), where they screened 30,000 compounds and found a link between, C*logP*, molecular weight and more favourable physiochemical

parameters.[118] The rule is aptly named 4/ 400 as it relates to those more favourable compounds having a C*logP* < 4 and a molecular weight < 400.[118] Another pharmaceutical company Pfizer investigated the *in vivo* tolerability of a group of 245 compounds and found a connection of C*logP* and topological polar surface area (TPSA) with a reduction of toxicity.[119] Those compounds with a C*logP* < 3 and TPSA > 75 were found to have a six fold reduction in toxicity, in comparison to those compounds with a C*logP* > 3 and TPSA < 75 where the reduction in toxicity was increased to 24 fold and as such this became the Pfizer 3/75 rule.[119] A major cause of toxicity can be attributed to target promiscuity; a screen of 2133 compounds found a direct correlation between C*logP* and increased chances of compound promiscuity.[120] Those compounds with a C*logP* < 3 had a reduced chance of being promiscuous whereas those with a C*logP* > 4 were found to be less specific towards their target and more promiscuous in nature.[120]

It has also been suggested that the higher the lipophilic nature of a compound, the more likely it is to be insoluble. The lipophilicity of a compound can be improved by reducing the number of aromatic rings, generally less than three is acceptable.[121] However if a compound is not lipophilic enough, then it may not be absorbed through the intestinal wall. Another way to improve the administration of a poorly soluble compound is through changes in the formulation, but by doing so the body will have to work much harder to metabolise and eventually excrete it. This would, therefore, result in the drugs needed for a longer duration of action and is an excellent example of the issues facing medicinal and computational chemists.

There is a fine balance needed for circumventing the issues associated with increasing lipophilicity in the search for improved drug potency. One method developed by medicinal and computational chemists to overcome these issues could be the use of a universally applicable index, the ligand efficiency. Ligand efficiency is defined as the energy of binding per non hydrogen atom and is very much dependent upon the size of the compound.[122] It is a direct correlation between the docking results and the physiochemical property of a ligand and can be useful for a more balanced selection of a subset of compounds.

1.4.5 Pareto Ranking

A balanced selection of a subset of compounds takes into account not only ADMET parameters but also all the scoring functions under a *Pareto* ranking system in order to determine ideal hit compounds with the best overall profile (potency and safety). The balanced selection or multi-objective optimisation is the linear combination of

multiple objectives into a single-objective function.[123] Furthermore it is the process to detect a 'pool' of optimal equivalent solutions that meet a wider range of attributes relevant to a problem.[123]

Research into multi-objective optimisation was conducted by *Pareto*, whose methodology was based on equivalent solutions, also known as non-dominated solutions being equally good as they represent numerous possible compromises among the objectives.[123] One solution can dominate another if it is better in at least one objective; the solutions are then ranked according to their dominance value.[113] *Pareto* selection enables many aspects of a compound to be assessed simultaneously. Once the solutions have been ranked in the *Pareto* format a number could be selected for synthesis and subsequent biophysical analysis.

A balanced compound selection can be achieved through the use of a number of different computational programs including Pipeline Pilot and Knime.[124,125]

Pipeline Pilot is a chemo-informatics tool developed by Accelrys that can be used to develop protocols to calculate the physiochemical properties of small molecules, data manipulation and computational filtering.[126] Knime uses a workflow methodology; it is based on the open source platform Eclipse and Java.[127] It was not originally intended as a chemo-informatics tool and is in fact used in many other disciplines that require data handling such as the finance industry.[126] However, its data mining background makes it a useful tool to use in conjunction with Pipeline Pilot to calculate to carry out a multi-objective assessment, such as the calculated physiochemical parameters, combined with the docking data for each individual compound to assess the overall profile.

.

1.5 Synthetic approaches to targeted PPIs

The challenges associated with modulating protein-protein interactions (PPIs) have meant that the area of high-throughput organic synthesis has become an established strategy in the development pipeline. Significant challenges in organic synthesis and our knowledge and understanding of protein-protein interactions have occurred in parallel to one another, with the development of structural biology techniques and specifically designed chemical libraries. As such the two techniques can be used together to establish methods for the chemical intervention of PPIs.[42] Research into the design and synthesis of small molecule inhibitors has been conducted for a number of key protein-protein interactions that are believed to play a vital role in many biological processes. Due to the central role of PPIs the number of interactions studied is increasing exponentially and some of the most interesting include, IL-2, HPVE2, ZipA, TNf, HDM2 and Bcl-X_L.[43] There are three chemical approaches for targeting these different PPIs using small molecules including peptide-based, natural product inspired and chemistry inspired/ diversity orientated synthesis as outlined below.

1.5.1 The peptide-based approach

The peptide-based approach involves the design and synthesis of small molecules that are able to mimic the secondary structure of proteins whether that be α-helix, β-turn or β-strand.[128–130] The peptide-based approach can be further subdivided into peptidomimetics and non-peptide small molecule structures, the first of which has been used extensively and has seen numerous successes, for example in targeting glycoprotein coupled kinase receptors.

Extensive investigations, include mimetics of peptides, peptide-turn and β-peptides to target PPIs for example those by Zhao *et al.*, Stigers *et al.* and Fletcher *et al*.[128,131–133] However as the main focus of this thesis is the design of small molecule modulators we shall discuss the design and synthesis of non-peptide small molecules. These compounds are able to mimic the secondary structure of a particular protein binding partner through the interactions of a particular motif.[42] Benzodiazepines were amongst the first classes of small molecules synthesised for the purpose of PPI modulation.[42] They are successful β-turn (**1.34**) and α-helical (**1.35**) mimetics, depending upon the orientation of their functional groups.[42]

Extensive investigations in this area have meant that there have been significant developments in synthetic strategies, particularly solid phase synthesis, more

specifically the modifications and linker chemistry domains where the developments of triazene linkers for post cleavage aza-Wittig reactions.[134,135]

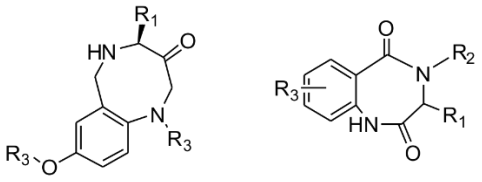
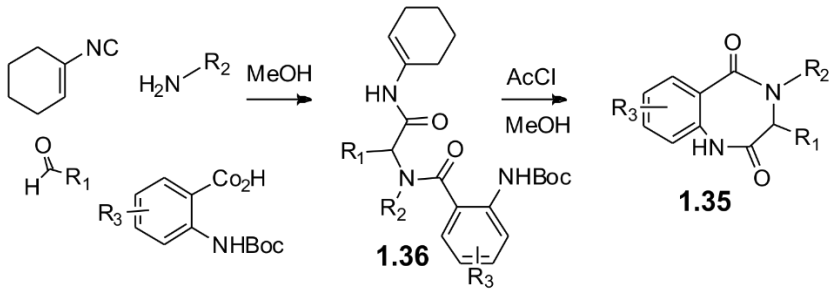


Figure 1.20- The benzodiazepine structures as β-turn and α-helix mimetics (adapted from [42]).

This class of compounds have been successful in targeting the p53 & HDM2 interaction, where an α-helix mimetic was developed and synthesised from a library of 22000 novel benzodiazepines.[136] The process involved using the Ugi-lactamization strategy in a two-step synthesis, the initial step involved combining four components in a condensation reaction (Scheme 1.6).[136] Subsequently, lactamisation of **1.36** involved acid catalysed cleavage of cyclohexenamide and simultaneous Boc deprotection to expose the anthranillic nitrogen to allow cyclisation.[136]



Scheme 1.6- The synthesis of P53 MDM2 inhibitors. The synthesis of the benzodiazepine library of 22,000 compounds using the Ugi-lactamisation strategy. The initial step involved combining four components in a condensation reaction, followed by the post condensation lactamisation of **1.36** which involved acid catalysed cleavage of cyclohexenamide and simultaneous Boc deprotection to expose the anthranillic nitrogen (adapted from [42]).

Hamilton *et al.* have extensively investigated α-helix mimetics that are based on rigid non-peptide scaffolds, such as terphenyl **1.37** and trispyridylamide **1.38** compounds (Figure 1.21).[77,137,138] Other rigid non-peptide scaffolds include biphenyl, indane, terpyridine, polycyclic ether, terephthalamides and trisbenzamides.[77,139–144] The functional groups displayed by these scaffolds adopt a particular orientation which is able to mimic the orientations of the α – R groups of the amino acid within an α helix.[42] Due to the issues associated with the solubility of these compounds, they have had limited use although the Hamilton scaffolds have been shown to be more suitable. An example of the biological uses of these scaffold types include the specific targeting of the Bcl-w Bak interaction.

Dömling *et al.* developed a diphenyl-imidazole scaffold using a multi component reaction in a one pot synthesis.[145] They generated 60 compounds to screen using cell proliferation assays and fluorescence polarisation assays which indicated three hit compounds for the intended target.[145]

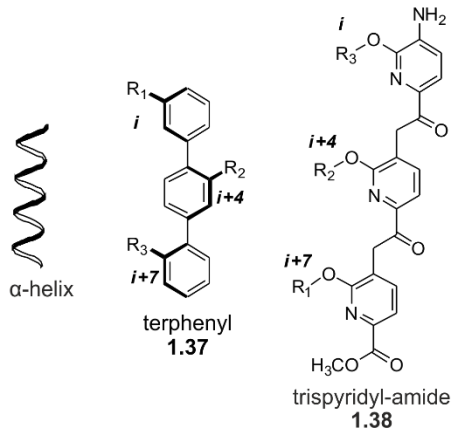


Figure 1.21- Hamilton's α-helix mimetics. The rigid scaffolds of the α-helix mimetics developed by Hamilton *et al.* The R groups are designed to mimic the amino acid side chains at positions i, i + 4 and i + 7 within an α-helix, with rigid terphenyl or tris pyridyl-amide core scaffolds maintaining the specific position along the mimetic.(adapted from [42]).

β-Turns play a key role in molecular recognition and the mimetics are often bicyclic with modified amide structures. There are several key scaffolds including benzodiazepines, diketopiperazines and isoindolines (Figure 1.22).[42] To identify modulators of the c-myc-max interaction, around 240 isoindoline scaffolds (120 amide derivatives and 120 acid derivatives) were prepared by Boger *et al*.[146,147] These bicyclic isoindoline-5,6-dicarboxylic acid templates contained three functionalised positions and were synthesised using solution based techniques.[146,148] The use of fluorescence resonance energy transfer (FRET) screening investigations led to a single hit compound that exhibited selectivity against the desired target with an IC_{50} of 20 μM.[149]

The bicyclic diketopiperazines are another rigid non-peptide scaffold that is able to mimic β-turns, synthesised using solid phase synthesis.[150] A high diversity of the different bicyclic fused and spiro structures can be introduced using different synthetic sequences.[151,152] An example of the ease of synthesis and diversity of compounds can be observed with the 5000 scaffolds prepared by Kahn *et al*., designed to modulate the interaction between β-catenin and cyclic AMP which would hence stop the TCF-mediated transcription processes.[153,154]

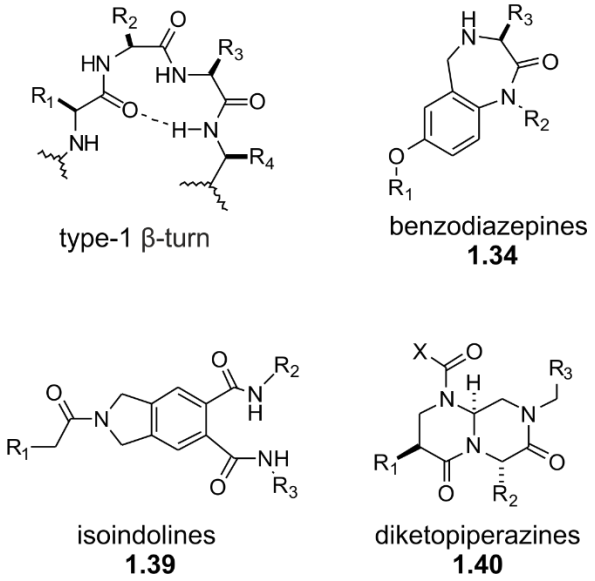


Figure 1.22- Small molecule compounds able to mimic a type-1 β-turn including benzodiazepines **1.34**, isoindolines **1.39** and diketopiperazines **1.40.** The R group substitutions of each small molecule compound can be varied to mimic the side chains of amino acids at the R1, R2, R3 and R4 positions on the type-1 β-turn (adapted from [147]).

1.5.2 Natural Product-inspired ligand design

Many drugs currently on the market have their structures derived from natural products. This may be due to their predisposition to interact with proteins from evolutionary development.[155,156] These are often large and complex molecules that derive their activity from PPI inhibition, contain specific functional groups responsible for their action. It is the challenge of synthetic medicinal chemists to recognise these important chemical features and develop inspired chemical libraries to aid the discovery of new PPI inhibitors.[157,158] There are a number of natural product inspired scaffolds that have been discovered as modulators of a number of PPIs including guanidine-tricycles **1.41**, spiroketals **1.42** and oxindoles **1.43.**

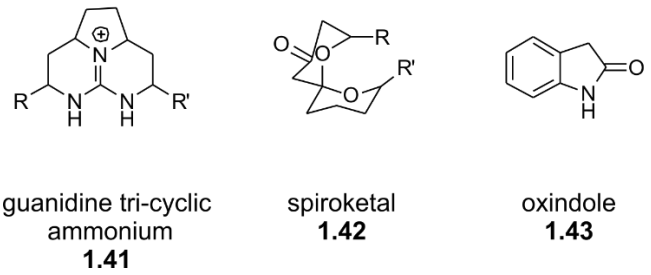


Figure 1.23- The core scaffolds of three natural-product inspired small molecule compounds developed from natural products and synthesised to target PPIs.

Tricyclic guanidine containing alkaloids such as batzelladines and crambescidines, both of which are marine natural products, were found to be PPI inhibitors. Batzelladines are able to disrupt CD4 interactions through either inhibition of binding or by inducing complex disassociation.[159–164] Both batzelladine and crambescidine analogues were found to disrupt a number of interactions related to

Nef complexes with p53, actin and p56lck.[162,165] Analogues of these complex scaffolds were synthesised using a number of complex total syntheses.[166]

Spiroketals are a motif that is common in many biologically important natural products that are structurally complex and have been shown to be active against a number of protein targets. Such spiroketal containing scaffolds include: altohyrtins (spongistatins) which are able to modulate microtubule formation bistramide A which suppresses cellular proliferation and didemnaketals A and B that target HIV-1.[167–172] Due to the diversity of their biological targets there has been keen interest in the development of libraries of spiroketal containing scaffolds for screening purposes.[173] A library containing a diverse set of compounds was developed by Porco *et al*.[173] They devised an efficient synthetic approach to an initial core spiroketal scaffold from an entiomerically pure epoxide.[173] The core scaffold was then diversified into 90 compounds using a three step synthesis to introduce a high level of structural complexity.[173]

There are a number of other groups that have applied different approaches to synthesising these scaffolds such as Lay *et al.* & Tan *et al*.[174–176] Finally spiro-oxindoles are another example of successful natural product like scaffolds that are found in numerous alkaloid natural products, including spirotryprostatin, pteropodine and alstonisine.[177] They have been implicated in inhibition of numerous biological systems, such as the p53 MDM2 interaction. Using the p53 MDM2 crystal structure, Wang developed a compound that was able to mimic the key side chain belonging to W23 of p53, with 3nM activity in the disruption of the p53 MDM2 interaction.[68,178–180]

There have been numerous libraries of spiro-oxindole like compounds developed using novel efficient synthetic approaches. Schreiber *et al.* employed a three-component strategy along with a split pool approach in a solid phase approach.[181] Using 16 spiro-oxindole core scaffolds they produced 3000 derivatives, using Sonogashira couplings, amidation reactions and *N*-acylations.[181] Following biological screening, the group further developed a hybrid library that contained 384 diverse compounds.[181,182] These scaffolds are of key interest to a number of researchers and have led to the development of numerous compound libraries through a variety of synthetic approaches.

1.5.3 Chemistry inspired through Diversity Orientated Synthesis (DOS)

Diversity orientated synthesis describes the synthesis of collections of structurally diverse small molecules, developed without any prior knowledge relating to the

structural requirements of a biological target.[42,183] The libraries are refined to contain molecules that have a better probability of surviving the “hit to lead” optimisation processes and this can be done in a number of ways.

One can invoke a number of structural requirements based on “the rule of five” developed by Lipinski *et al.* in 1997, “the rule of three” developed by Oprea and finally lead likeness.[184–186] However, compared to compounds aimed at targeting other biological targets such as enzymes, most current PPI inhibitors are essentially not “drug- like”. They tend to have high molecular weights, are complex and lipophilic, therefore targeting these interactions will require screening of diverse compound structures obtained by diverse orientated synthesis.[187–191] The compositions of each library and the synthetic approaches used to generate the compounds may differ somewhat.

One such library was developed by Boger *et al.*, who believed that due to the structural prerequisites of the large interaction sites of many protein-protein interactions, the compounds to target them would have a greater chance of doing so, if they themselves were much larger.[192,193] The initial library of compounds had a high degree of molecular complexity and the synthetic approach was fast and efficient. It involved iminodiacetic acid diamides **1.44** (Figure 1.24) which were subsequently dimerized using alkene metathesis to produce solutions (pools) containing a mixture of *cis* and *trans* alkenes **1.45**.[192,193] By doing so, the initial screening procedures were made more efficient at screening higher numbers of compounds against their target. However, the disadvantage to this process was the tricky deconstruction of the results even if the number of compounds in each pool were low. It was also found that the diversity of SAR contributions was low, this was because the contributions of less potent compounds were not taken into consideration, instead the most potent compounds of the pools dominated.[192,193]

To overcome this problem the group designed a second library which eliminated the possibility of alkene isomers by not using alkene metathesis to dimerize the diamides, instead a more rigid di-acyl linker was introduced **1.46** (Figure 1.24).[192–195]

The secondary library contained 60 pools of 10 compounds which were screened against the protein complex of mmp2 and αvβ3. Of the 60 pools three were believed to contain hits and so the 30 individual compounds were synthesised and purified, furthermore the most potent single compound of the 30 was selected.[192–195] This compound was diversified into a further 77 diverse analogues, the most successful of

which was water soluble and found to have a high degree of potency towards the inhibition of angiogenesis.[192–195]

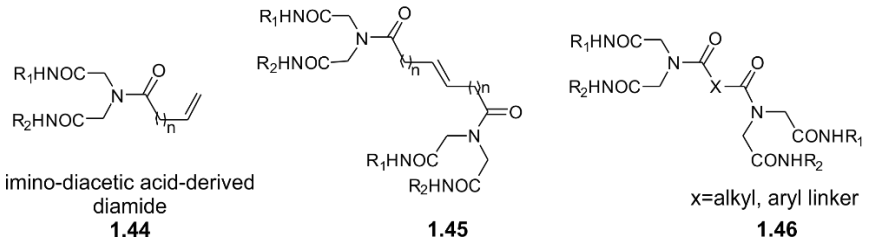


Figure 1.24- Small molecule scaffolds from the Boger diversity oriented synthesis library, the initial imino-diacetic acid-derived diamide **1.44** undergoes dimerization using alkene metathesis to form **1.45**. Optimization of the core scaffold of the library involves replacement of the flexible alkene linker with a more rigid alkyl or aryl one **1.46** (adapted from [42]).

There are many other examples of different DOS libraries such as those synthesised by Janda *et al.* to contain a central flat aromatic core, utilising quinoline **1.47** and napthalene **1.48** structures (Figure 1.25).[196,197] Schreiber *et al.* used solid support split pool synthesis to develop a number of different libraries containing diverse dioxolanes, biaryl macrocyclic ethers and amides (Figure 1.25).[198,199] Finally Bienayme used an efficient alternative to multistep synthesis, the multicomponent reactions (MCRs) were key in the development of the library of 8160 compounds from which an inhibitor of the VEGF/ NRP1 interaction was discovered.[200]

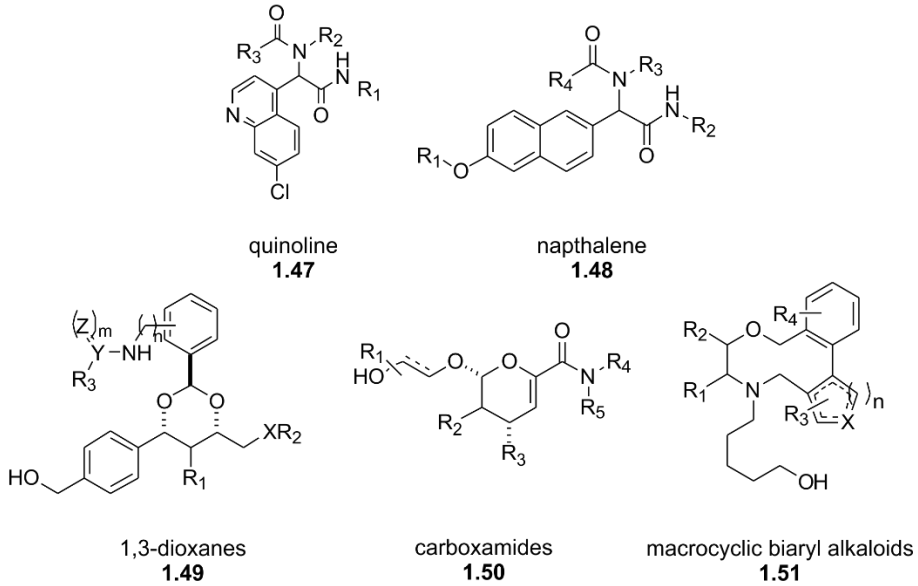


Figure 1.25- Other core scaffolds from different diversity oriented synthesis libraries developed by Janda *et al.* and Schreiber *et al.* to target PPIs using diverse small molecules(adapted from [42]).

Thus three different approaches: peptide based, natural product inspired and chemistry inspired/ DOS, can be used to develop compound libraries that can be screened against a biological target. They also describe the high through-put nature of the chemical synthesis, in order to produce a large number of compounds to give a greater chance of success against their targets. They utilise a variety of synthetic

procedures in peptide coupling, with a wide range of chemical scaffolds. The development of these three approaches has occurred over a number of decades and often followed a structure-based drug design, using computational chemistry and structural biology to aid the synthesis of more potent compounds. In each case there is often described an initial single compound from which a library of small molecules is designed following synthetic optimisation.

Having discussed synthesis of molecules an assessment must be made of their PPI modulation *via* biophysical techniques.

1.6 Introduction to biophysical techniques

In any structure-based drug design project for modulators of protein-protein interactions (PPIs) it is important to understand how the synthesised ligand interacts with the protein of interest, this will enable optimisation of the ligand in future synthesis.

To do this there are a number of biophysical techniques which can be employed such as; X-ray crystallography; fluorescence polarisation (FP); nuclear magnetic resonance (NMR) spectroscopy; isothermal titration calorimetry (ITC) and intrinsic fluorescence spectrophotometry. These techniques have been used extensively in many successful drug design projects that aim to target protein-protein interactions. It is the last three that are used in this thesis and shall be discussed briefly here, followed by the applications of those techniques in PPI investigations.

1.6.1 Protein Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is able to determine characteristic properties of molecules generally in solution. Such properties include chemical and electronic information as well as more in depth three-dimensional information on structure and conformation. It is a technique that has been used for the characteristic determination of small molecules and is also a key biophysical technique for structural analysis of larger macromolecules such as proteins.[201]

The phenomenon of NMR works under the principle that all matter is composed of different atoms which, are in turn made up of nuclei and electrons.[202] When exposed to a magnetic field some of the nuclei align themselves to the applied field. Furthermore, each atomic nuclei have four properties including; mass, electronic charge, nuclear spin and magnetism.[203] Nuclear spin can be observed as a frequency *via* NMR; the specific nuclei can be observed dependent on their atomic properties.

When the atomic quantum number (*I*) = ½ the atoms may be observed by NMR. Thus for NMR spectroscopy of biomolecules such as proteins, the most important nuclei are those which possess the characteristic *I* = ½ including ^{1}H, ^{13}C and ^{15}N; their properties can be found in table 1.1.

T**able 1.1- The nuclei of interest's characteristic properties.** Including; *I* the nuclear spin angular momentum, the natural abundance of each isotope measured as a percentage and γ the gyromagnetic ratio.[203]

Nuclei	*I*	Natural Abundance (%)	γ $(Ts)^{-1}$
^{1}H	½	99.9	2.6752×10^{8}
^{13}C	½	1.07	6.728×10^{7}
^{15}N	½	0.37	-2.73×10^{7}

As naturally occurring proteins are composed of around 20 different amino acids which contain protons in different chemical environments, small local changes in magnetic fields around an atom due to its particular environment alters the frequency at which the protons come into resonance. The different frequencies of the hydrogens results in different chemical shifts of those nuclei that are the same, but which are surrounded by a different chemical environment e.g. the protons on two alanine side chains in different chemical environments will differ.

A chemical shift (δ) is defined as a specific resonance of a specific nucleus relative to reference, for ^{1}H and ^{13}C; this reference is usually tetramethylsilane (TMS) and the units of this values is given in parts per million (ppm).[203] The different chemical shifts of a protein are determined by a number of factors such as if the hydrogens are in an aromatic environment (i.e. residues Phe, Tyr, Trp or His).[204] Those in aliphatic regions include Leu, Ile, and Lys to name but a few.[204] The methyl groups of the different amino acids occur at specific ^{1}H ^{13}C chemical shifts and the same can be said for some of the ^{1}H ^{15}N amide NH peaks.[204] These differences are observable in different spectra, for example a simple one dimension (1-D) ^{1}H experiment (Figure 1.26 C), a 1-D spectrum is simply the chemical shifts of the ^{1}H nuclei within a molecule.

There are also two and three dimensional (2-D, 3-D) experiments and these take into account the transfer of magnetisation from the ^{1}H nucleus to the connected atom. Such 2-D experiments include ^{1}H ^{13}C and ^{1}H ^{15}N heteronuclear single quantum correlation (HSQC) (Figure 1.26, a and b). Heteronuclear experiments show a one bond correlation between nuclei, in the case of a ^{1}H ^{15}N HSQC spectra, the number of peaks represent the number of amino acids within the protein excluding proline residues which are unobservable.

In the case of 2-D experiments such as a ^{1}H ^{15}N HSQC experiment the magnetisation is transferred from the hydrogen onto the attached ^{15}N nuclei, the chemical shift is then derived from the nitrogen; magnetisation is next transferred back to the hydrogen where it is then detected.[203]

These experiments together provide a comprehensive finger print of all the nuclei within the protein. This is important in the determination of the proteins secondary structure. Both 2-D and 3-D experiments are used in the resonance assignment of a protein spectrum. Changes in the chemical environment of amino acids within a protein can occur due to interactions with a ligand or natural binding partner. These changes can be monitored using a number of different experiments such as 2-D HSQC's, 1-D saturation transfer difference (STD) and 1-D waterLOGSY experiments (Chapter 4.1.1).

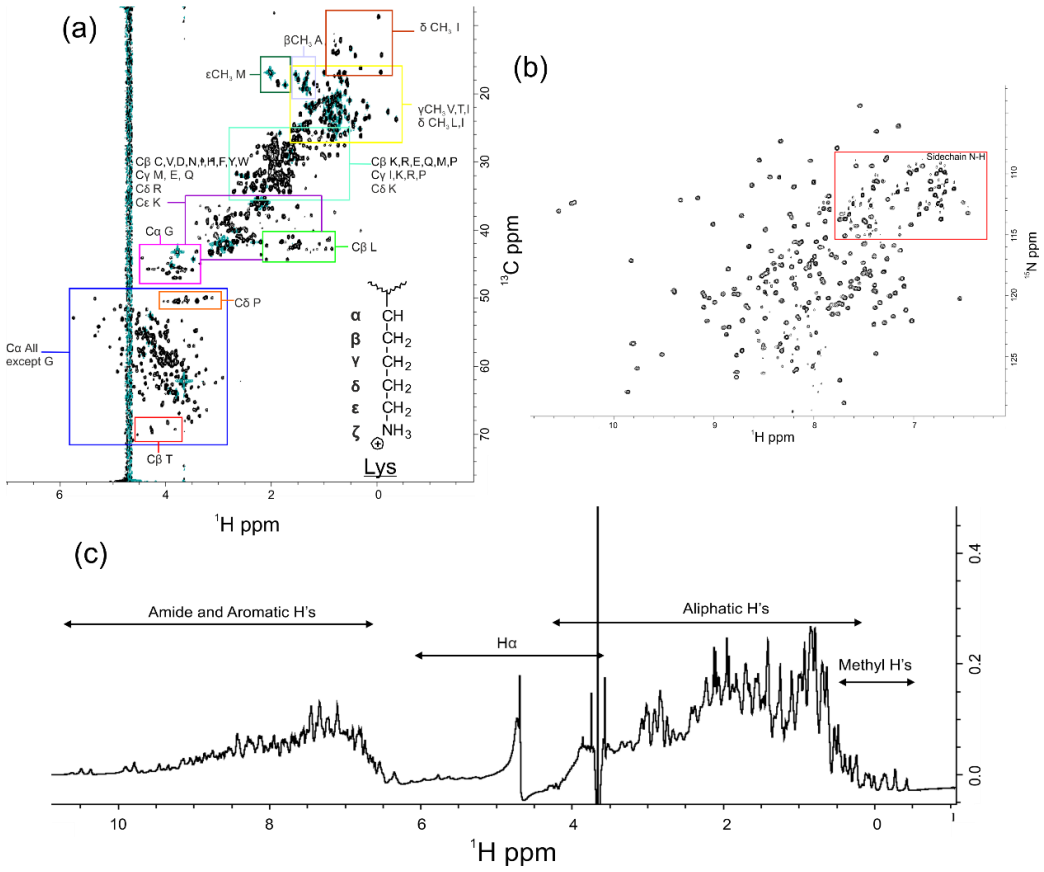


Figure 1.26- Example 1-D and 2-D spectra. (a) A ^{1}H ^{13}C HSQC spectra of NCS1 (500μM) with annotations relating to the assignment regions of specific carbon or hydrogen environments. An example amino acid Lysine can be seen with the nomenclature annotated. **(b)** A ^{1}H ^{15}N HSQC spectra of NCS1 (500μM) each peak in the spectra relates to a specific backbone NH amide of a particular amino acid within the protein, the number of peaks within this spectrum provides a rough idea of the number of amino acids within a protein, excluding prolines. The area highlighted in the red box relates the side chain NH's of residues including Arg, Gln and Asn. **(c)** A 1-D ^{1}H spectra of NCS1 (250μM) with the specific regions highlighted. (All data collected in this figure was collected at the University of Liverpool and used for the purposes of this research project).

The spectrometers used to collect the NMR data in this thesis include either a Bruker AVANCE II+ 600 MHz Ultrashield or 800 MHz US^2 spectrometer, Bruker AXS Inc., Madison, Wisconsin, USA, each spectrometer is equipped with a 5 mm triple-resonance cryoprobe. The increase in field strength between 600 MHz and 800 MHz results in an increase in the resolution, the use of a cryoprobe increases the sensitivity of the ^{1}H, ^{15}N and ^{13}C detection.

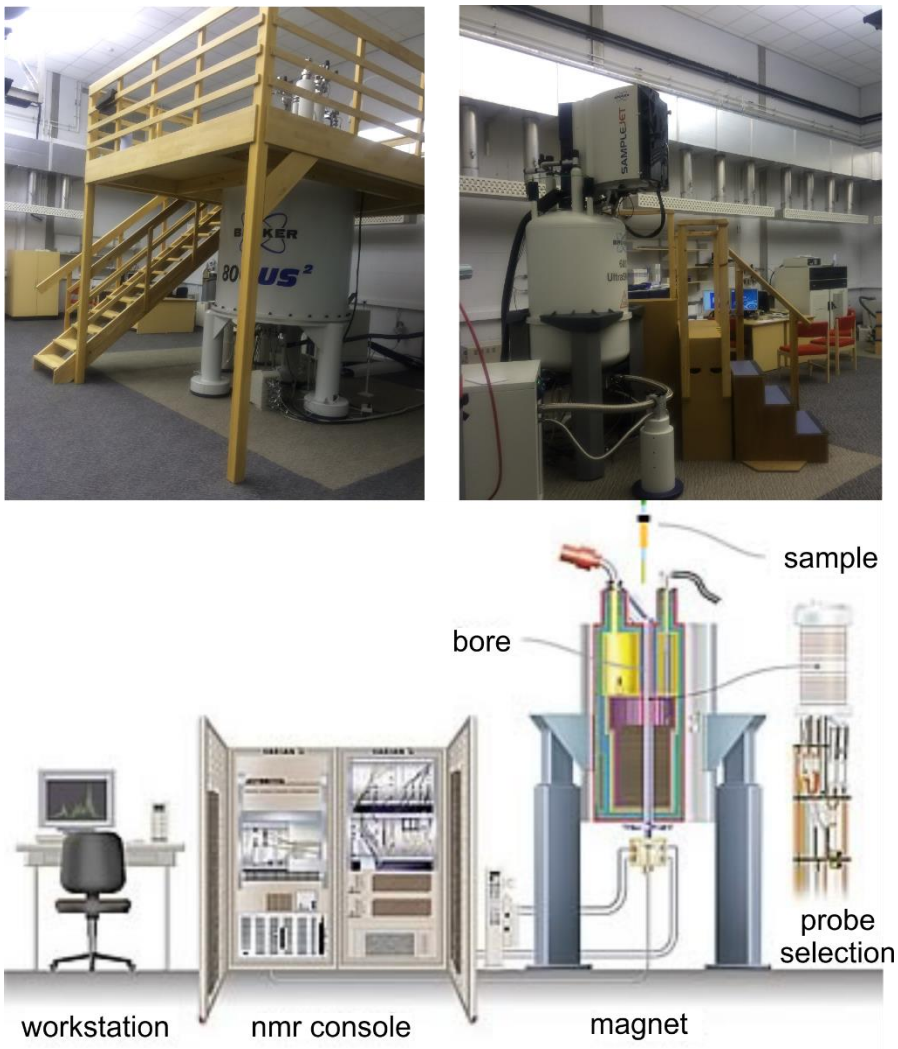


Figure 1.27- The spectrometers employed at the University of Liverpool Centre for Structural Biology. The Bruker AVANCE II+ 800 MHz and 600 MHz spectrometers, alongside a basic schematic of the setup of a spectrometer and processing equipment (Schematic adapted from www.varianinc.com).

1.6.2 Isothermal titration Calorimetry

The term calorimetry is the description of a change in heat of a system that occurs as a result of a physical process. Experimentally this change in heat is measured using a calorimeter.[205] The technique has early origins and has been used extensively over the centuries to investigate numerous physical processes such as measuring the heat capacity of water carried out by Black in the late 1700's.[206] Calorimetric measurements can be used to explore systems in equilibria through a titration method. Original chemical applications studied the exploration of acid- base equilibria and metal ion complexation reactions.[207,208] It was not until later in 1989 that MicroCal developed a calorimeter to investigate the biological equilibria between Cytidine 2'-monophosphate (2'CMP) with the active site of ribonuclease.[209] Isothermal titration calorimetry is the measure of a change in heat of a physical process, with the sample of interest and reference sample kept at the same constant temperature during the experiment, hence its term isothermal.[210]

ITC is able to provide a detailed thermodynamic characterisation of the interaction between a protein and a ligand or secondary binding peptide/ protein, such thermodynamic properties include enthalpy (ΔH) and entropy (ΔS), free energy (ΔG) and binding affinity (K_D).[211] Each thermodynamic property is related to the other using the following equations (1.1 and 1.2), ΔG and K_D are derived from the measured thermodynamic properties, with ΔH and ΔS measured experimentally. As a technique employed to investigate the interactions of biological systems it can be used alone; however generally it is used in conjunction with other complementary biophysical techniques such as NMR spectroscopy, tryptophan fluorescence and surface plasmon resonance (SPR).

$$\Delta G = \Delta H - T\Delta S \qquad 1.1$$

$$\Delta G = -RTlnK_D \qquad 1.2$$

The ITC200 Microcalorimeter (GE Healthcare, Figure 1.28 a) was the calorimeter used for the ITC experiments in this thesis, a schematic of the calorimeter (Figure 1.28 b) highlights the three key components. The reference cell (blue) which contains either a sample of water or the corresponding buffer, the sample cell (magenta) contains the titrand into which the computer controlled syringe (green) injects and stirs the titrant. Both the reference cell and sample cell are maintained at a constant single temperature, upon injection of the titrant into the titrand any change in heat from the sample cell is measured and the data is then read by the computer for analysis.

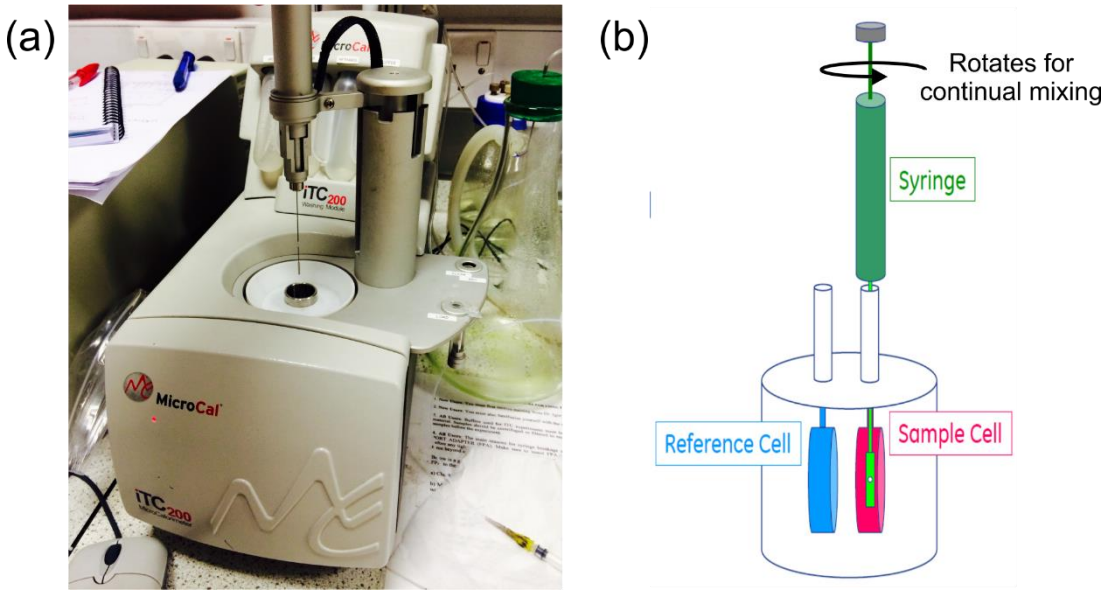


Figure 1.28- **(a)** The ITC200 Microcalorimeter (GE Healthcare) of The University of Liverpool Centre for Structural Biology on which all experimental ITC within this thesis was conducted. **(b)** A schematic representation of an ITC calorimeter with the three key components indicated. The reference cell (blue) contains either a sample of water or buffer, the sample cell (magenta) contains the titrand into which the computer controlled syringe (green) injects the titrant. As the experiment proceeds the syringe rotates for continual mixing and both the reference cell and sample cell are maintained at a constant single temperature. Upon injection of the titrant into the titrand the change in heat from the sample cell is measured and the data is then read by the computer, (Image adapted from GE Healthcare MicroCal calorimeters manual).

The raw data obtained from each experiment is analysed in Origin7 (OriginLab, Northampton, MA) using a curve fitting process in which a non-linear regression procedure is used to apply a specific mathematical model to the data (equation 1.3).[212] It works following an iterative process to minimise the chi-square (chi^2) values and obtain best fit parameters for the binding parameters.[213,214] The fitting process involves, (a) initial estimates by the Origin software of the values for n, K and ΔH and (b) calculation of ΔQ (i) for each individual injection, these values are subsequently compared with the heat measured for the corresponding experimental injection. Standard Marquardt methods (c) are used to improve the values of n, *K* and ΔH by continuous iteration of the procedure (a, b and c) until there is no improvement in the values.

The analysis software is able to take the raw data and create an isotherm (Figure 1.29), this is a plot of the change in heat (Kcal) as a function of the ratio of ligands concentration: protein concentration. Subsequent analysis of the isotherm is able to minimise the chi-square values and calculate the 'best fit' values of the following binding parameters: n, the stoichiometry, K_D the binding affinity, ΔH the change in enthalpy and ΔS the change in entropy. The enthalpic contributions (ΔH) from any given experiment relate to the change in intermolecular bonds/ interactions, whether

this be formation or destruction of hydrogen bonding or van der Waals interactions.[211,213] The entropic contributions (ΔS) of a reaction come from the change in solvation of the apolar protein surface, which could be the burial of hydrophobic residues or the release of water molecules.[211,213] For an ideal interaction the thermodynamic profile would be as follows: the enthalpy ΔH would be large and negative, the component TΔS would be small and positive thus resulting in a large and negative ΔG which is indicative of a tight binding interaction.[211,213] Further details of the ITC methodology are discussed in the relevant Chapter 2.3.5.

$$\Delta Q(i) = Qi + \frac{dVi}{Vo}\left[\frac{Q(i)+Q(i-1)}{2}\right] - Q(i-1) \qquad 1.3$$

V_o= active cell volume

V_i= injection volume

Q= total heat content in V_o

ΔQ_i= heat released from the i^{th} injection

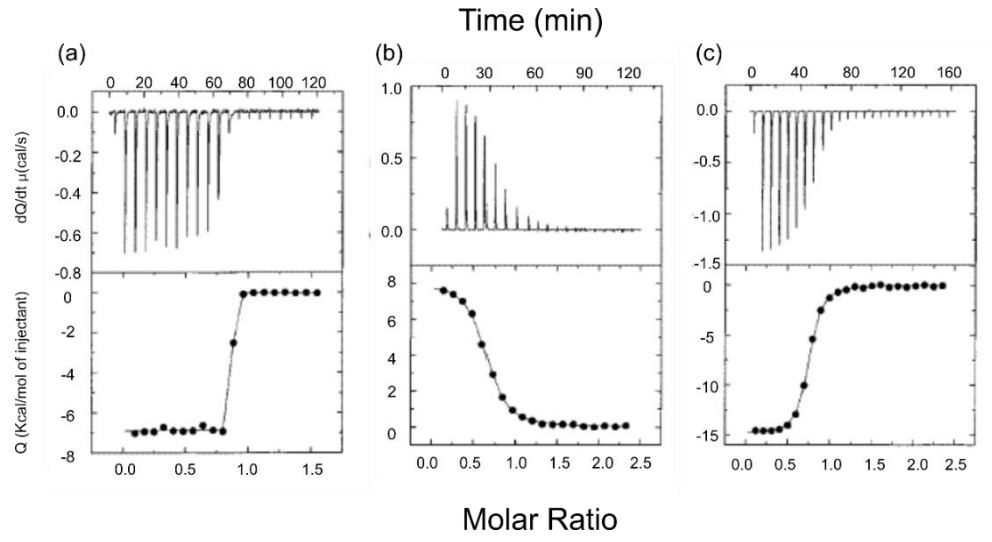


Figure 1.29- Examples of different isotherms (top) and resultant curves (bottom) acquired on a microcal200 in relation to the studies of the interaction of HIV-1-Protease with **(a)** Amprenavir **(b)** Acetyl Pepstatin and **(c)** Amprenavir and Acetyl Pepstatin (adapted from [210] and [215]).

ITC has been used in a number of biological investigations such as the binding affinity and stoichiometry between two proteins and between proteins, peptides or small molecules. However, as with any other technique it is not without challenges; it does have upper and lower limits with regards to binding affinities, which means that only those interactions within the range of 10 nM and 1 mM can be assessed effectively. In order to study interactions outside of this range competition experiments can be used for very strong or very weak binders. This does mean that to obtain good data,

planning the experimental conditions has to be extremely thorough. To do so requires a knowledge of an estimated binding constant, this enables the concentration of both the titrant and titrand to be calculated. Other requirements that should be considered include the pH, temperature and buffer conditions of both the protein and the ligand to avoid any mismatch that could result in artefact heat changes.

Despite these initial requirements it is still an extremely useful tool in the structure-based drug design processes. As changes to the instrument occur over time, making it more sensitive to those weaker interactions then its use will become even more prevalent.

1.6.3 Fluorescence Spectrophotometry

The emission of light from any given surface is known as luminescence and can be divided into two categories, fluorescence and phosphorescence, depending on the excited state.[216] Here though we shall discuss fluorescence of which there are many applications for biological investigations such as, fluorescence resonance energy transfer (FRET), fluorescence polarisation (FP) or anisotropy (FA).

Fluorescence polarisation assays are used widely in biological screening of small molecule interactions with proteins, it is based on the observation that when excited with plane polarised light, fluorescent molecules in solution will emit the light back to a specific plane so long as they remain stationary during the excitation phase.[217,218] However, most molecules do not stay stationary; they may rotate or tumble which has an effect on the planes into which the emission of light will occur.[217] Generally FP experiments are real-time experiments that are conducted in solution; they require the use of a fluorophore, and are very sensitive in nature.[217]

Fluorescence anisotropy can be used in a similar manner to FP, however it describes the phenomenon whereby the light emitted by the fluorophore after excitation is has different intensities along different axes of polarisation.[219] Fluorescence anisotropy has a number of applications and has been used to study the interaction of NCS1 with D2R peptide.[220]

Fluorescence resonance energy transfer (FRET), is the process of the transfer of energy from an excited fluorophore to another.[221] It is useful in probing protein interactions as the transfer of energy will only occur when both the donor and acceptor fluorophores are in close proximity to one another, this is around 1 - 10 nm.[221]

There is another fluorescence technique which does not rely on any external sources of fluorescence, intrinsic protein fluorescence. In particular intrinsic protein

fluorescence has been used for the investigation of the binding interactions of a protein and more details of protein structure. Intrinsic protein fluorescence exploits the naturally fluorescent aromatic sidechains of the amino acids tryptophan, tyrosine and phenylalanine within the protein itself, rather than a fluorescence label or dye.[216]

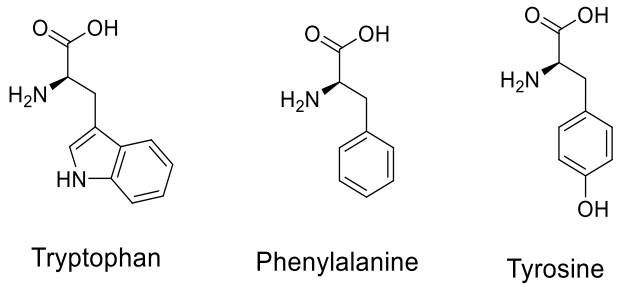


Figure 1.30- Chemical representations of the three aromatic sidechains of the amino acid residues Tryptophan, phenylalanine and tyrosine. Of which tryptophan is the most powerful fluorescence tool as it is stronger in emission than tyrosine or phenylalanine and is extremely sensitive to the local environment that surrounds the indole sidechain.

For a particular type of fluorescence the sample is excited at a specific wavelength (λ) of light relating to that residue, for example in tryptophan fluorescence the excitation wavelength is 285 nm and the emission wavelength is between 300 - 350 nm. Tryptophan fluorescence is highly sensitive to the local environment that surrounds the tryptophan's indole sidechain, hence tryptophan fluorescence can be used to detect conformational changes of a protein at low concentration.[216,222] These changes in conformation present themselves in two ways. A blue shift in the fluorescence is observed in those circumstances where the tryptophan residues become more buried into the hydrophobic surface upon interaction of the protein with a particular ligand or interacting protein.[216] A red shift in the tryptophan fluorescence occurs when the tryptophan residues become more exposed and this is a possible indication of protein unfolding.[216] Both the red shift and blue shift present themselves in the following manner, a blue shift is a shift to the left of the native fluorescence i.e. less than 300 nm, a red shift is a shift to the right meaning the emission spectra would occur above 350 nm.

It is a very useful method in detecting binding of a protein and ligand if the protein of interest is known to contain tryptophan residues in the binding site. In this case a possible interaction between the protein and the ligand may be observed in the quenching or enhancing of the fluorescence emission.[223] Not only is this technique able to determine structural changes in the protein upon ligand binding but it is also used to determine the binding affinity of the interaction.[223] The tryptophan fluorescence data collected for this thesis was collected on a Cary Eclipse

Fluorescence spectrophotometer (Agilent technologies), the raw data was analysed in Origin7 (OriginLab, Northampton, MA).

1.6.4 The applications of biophysical techniques: theory into practice

With the theory behind each of the biophysical techniques briefly presented, it is important to put their uses into the context of a drug design project.

Structure-based drug design revolves around the combined approach employing multiple disciplines; computational chemistry with synthetic chemistry and biophysical analysis, in order to determine hit compounds. The biophysical techniques have been used extensively in many successful drug design projects of protein-protein interactions. There have also been a number of instances where biophysical analysis has been used to confirm computational predictions in proof of principle situations.

An example of this can be seen in the work conducted by Hamilton *et al.* and the development of some terphenyl derivatives. A number of ^{1}H ^{15}N HSQC NMR experiments were used to confirm the docking poses of **1.52** with Bcl-xL. Identification of the amino acid residues of Bcl-xL affected by the ligand meant they were able to determine that **1.52** was interacting within in the Bak BH3 binding domain (Figure 1.31).[76] By combining the two disciplines, computational docking with biophysical analysis, the group were able to definitively identify the binding pose of **1.52** within Bcl-xL and hence use this to develop a second generation of scaffolds.

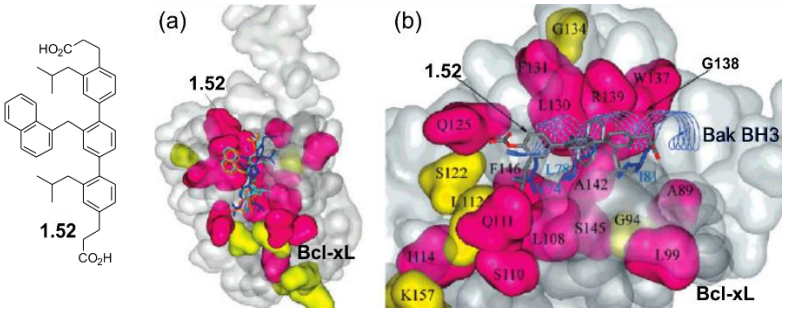


Figure 1.31- A comparison between the molecular docking of a terphenyl compound developed by Hamilton *et al*. and the ^{15}N HSQC results (adapted from [76])**.** The amino acid residues of Bcl-xL that are effected by the addition of compound **1.52** are coloured in yellow and magenta. **(a)** The top three binding poses of compound **1.52**. **(b)** An overlay of the best ranked binding pose of **1.52** with the Bak BH3 peptide (light blue), the three key sidechains of the amino acids Val74, Leu78 and Ile81 in the i, i + 4 and i + 7 positions.

With the second generation of compounds developed, the terephthalamides, Hamilton *et al.* used fluorescence polarisation alongside NMR spectroscopy to analyse the binding interactions of a number of the scaffolds.[77] The fluorescence polarisation competition experiments indicated that the terephthalamide compound **1.53,** was able to displace the Bak peptide from the Bcl-xL surface.[77] To further investigate this apparent competitive interaction, ^{1}H, ^{15}N-HSQC experiments were

employed which confirmed **1.53** was mimicking the helix of the Bak BH3 peptide (Figure 1.32).[77]

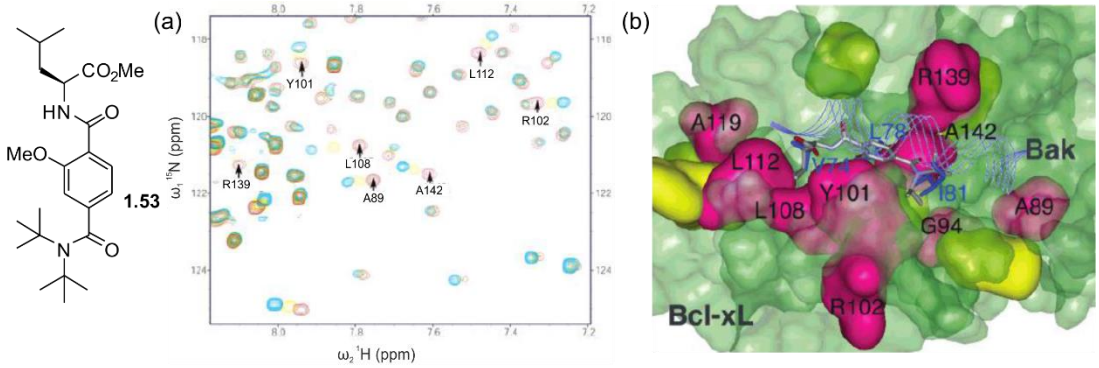


Figure 1.32- NMR results of terephthalamide compounds. The chemical structure of the terephthalamide **1.53**. **(a)** ^{1}H, ^{15}N-HSQC of Bcl-xL 200 mM without ligand (blue), overlaid with Bcl-xL in the presence of **1.53** (red), the residues that are severely affected can be seen annotated on the spectra. **(b)** The crystal structure of the Bcl-xL surface with the effected residues highlighted, the top ranked binding pose of **1.53** can be seen superimposed with the Bak peptide (adapted from [77]).

The final example of how biophysical techniques can be employed to analyse the binding interactions of designed compounds, can be observed in the development of the benzodiazepine scaffolds. Again Hamilton *et al.* used a combination of techniques to explore the binding interactions of a number of his designed inhibitors. As before, a competitive fluorescence polarisation assay was one of the initial techniques used to determine affinity of a number of compounds. Each compound was titrated into the fluorescent Bak peptide bound to Bcl-xL (Figure 1.33 (a)).[75] The most potent compound **1.54** was found to have a *Ki* = 2.4 μM and the least potent **1.55** *Ki* = > 500 μM (Figure 1.33 (a)).[75] Subsequently the thermodynamics of these binding interactions were investigated using ITC, **1.54** was found to have favourable enthalpic contributions, however the entropic values were small and negative and hence unfavourable for a tight interaction.[75] This could then be compared to the ITC curve of **1.55,** which supported its poor potency previously determined by fluorescence polarisation.[75] Finally ^{1}H, ^{15}N HSQC experiments were used to determine the region of Bcl-xL effected by the ligand.[75] This final experiment was able to confirm the competitive nature of ligand **1.54** as it indicated the residues effected as being in the Bcl-xL binding region.[75]

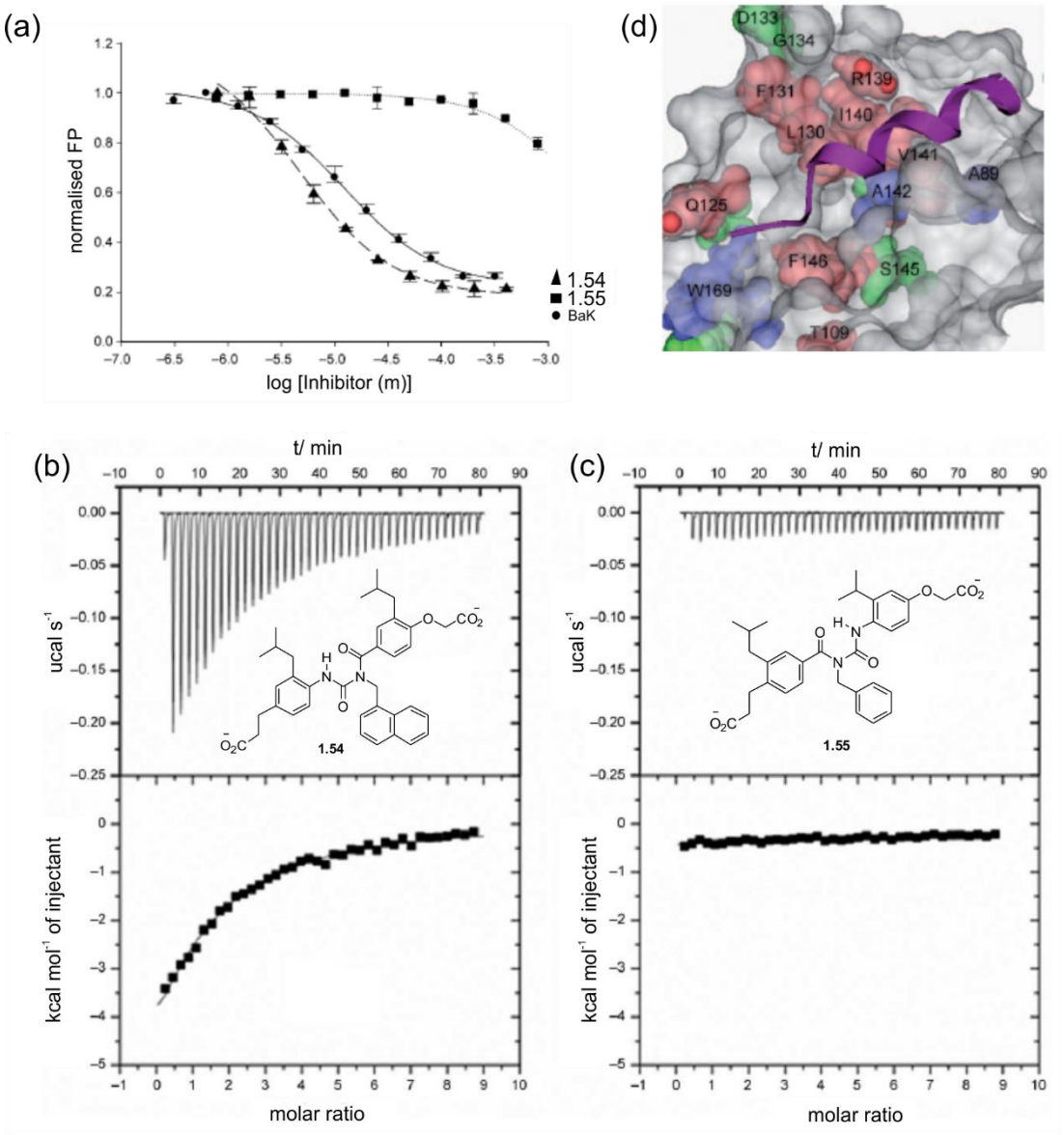


Figure 1.33– Bio-physical results of benzodiazepine compounds obtained by Hamilton *et al.* (a) Fluorescence polarisation titration curves obtained for **1.54** and **1.55. (b)** Normalised experimental ITC curve obtained from a titration of ligand **1.54** into Bcl-xL. **(c)** Normalised experimental ITC curve obtained from a titration of ligand **1.55** into Bcl-xL. **(d)** Crystal structure of Bcl-xL and Bak BH3 helix, with the residues effected by NMR studies of **1.54** highlighted (adapted from [75]).

1.7 Summary

This introductory Chapter has detailed the challenge and previous approached methods for targeting protein-protein interactions and reviewed the different approaches to discovering quality hit compounds as seen with the MDM2/ p53, and Bcl-xL/ Bak BH3 interactions.

It has highlighted the pharmacophore-based virtual screening approach as the vital first step in the holistic structure-based design process and gave a detailed overview of the different computational processes that can be applied to aid the selection of hit compounds to synthesise and test. Also considered here is the range of different synthetic organic chemistry protocols applied to synthesise some of the well-known scaffolds, from which successful PPI inhibitors were developed.

The basis of the biophysical techniques used within this thesis has been introduced to provide an understanding as to the meaning of their results and their purpose within the grounds of a structure-based design protocol. These three techniques will be used to design small molecule inhibitors of a previously unsuccessfully targeted protein-protein interaction such as NCS1 and D2R peptide. Furthermore the research details the structural characteristics of the individual proteins NCS1 and D2, alongside that of the complex, the physiological implications of the interaction and hence presents it as a valid target of interest for research.

1.8 Project Aims

Despite the evidence of the importance of the NCS1 D2R peptide interaction as a therapeutic target, the research into small molecule modulators of the NCS1 D2R peptide interaction had not been carried out. Therefore, research focussed on the discovery and development of novel small molecule inhibitors through a structure-based design protocol is of great interest.

The aim of this project is to discover novel small molecule inhibitors of the protein-protein interaction between NCS1 and the D2R peptide, using a structure-based design strategy.

In order to achieve this the project can be subdivided into four sections:

1. Development of a computational workflow involving a pharmacophore-based modelling, to screen a library of compounds and hence make a selection of ligands to synthesise (Chapter 2.1 and 3.1).

2. Design and conduct efficient synthetic routes to a selected number of small molecule ligands (Chapter 2.2 and 3.2).

3. Analyse the interactions of the synthesised compounds with NCS1 using a variety of structural and biophysical techniques. From the experimental measurement then optimise through molecular design, the next generation of compounds, for synthesis and testing (Chapter 2.3 and 3.3).

4. Investigation of a complementary approach through the use of a fragment library of compounds and apply a similar computational workflow processes to select compounds to purchase and screen against NCS1 using NMR-based methods (Chapter 4).

The techniques employed during this research include;

Computational Chemistry- pharmacophore approach to the structure-based drug design process, docking, scoring and ranking of compounds. Programmes used were: ZINCPharmer, GOLD, Pipeline Pilot, Knime, Spartan and PyMOL. [92,99,124,125,224,225]

Synthetic Organic Chemistry- Using synthetic organic chemistry methods to synthesise the selected inhibitors in an efficient manner.

Protein Production- Following a protocol previously published protocol to produce unlabelled and isotopically labelled ^{1}H ^{15}N and ^{13}C ^{15}N NCS1 for the requirements of fluorescence assays, ITC and NMR analysis.

Intrinsic Tryptophan Fluorescence- Employed as an initial screen to determine if there was an interaction between the ligands and the unlabelled NCS1.

NMR Spectroscopy- Utilised as the method to determine the sidechain and backbone assignment of ^{13}C ^{15}N NCS1 and the location of any interactions between NCS1 and the inhibitors. Use two-dimensional ^{1}H ^{15}N and ^{13}C ^{15}N HSQC experiments alongside some one-dimensional experiments such as saturation transfer difference (STD) and waterLOGSY (see Experimental Chapter 6 for further details) to screen for binding.

Isothermal Titration Calorimetry (ITC) - The technique used to determine the binding affinity of any NCS1 ligand interaction and any thermodynamic parameters associated with any binding interaction.

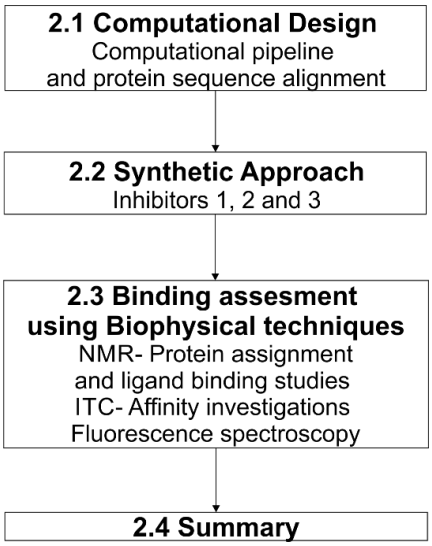
Chapter 2

Structure-based drug design: The First Generation

2 Structure-based drug design: The First Generation

The research detailed in this chapter, describes the structure-based molecular design, organic synthesis and biophysical assessment of a series of first generation compounds. An outline of the research discussed in this chapter can be found below (Scheme 2.1). It incorporates details of the computational design, synthetic approach and biophysical analysis of three inhibitors designed to target the protein-protein interaction between neuronal calcium sensor 1 (NCS1) and the dopamine receptor 2 (D2R).

The biological implications of the interaction between NCS1 and D2R include bi-polar and schizophrenia disorders, making it an important therapeutic target (a greater discussion of the implications of both NCS1 and D2R can be found in the introduction Chapter 1.1.1). At the initiation of this project the crystal structure of the NCS1 D2R complex was unavailable. However, the sequence of NCS1 was known and a sequence alignment between the yeast NCS orthologue frequenin (frq1) and rat NCS1, indicated that the two NCS proteins share a 60% sequence identity and an overall sequence similarity of 75%. Therefore frq1 and its binding partner phosphatidylinositol 4-kinase (pik1) could be used as the model system for the computational design processes. A sequence alignment between the two NCS proteins can be found in Chapter 2.1.6 and further details on the frq1 pik1 interaction can be found in the introduction Chapter 1.1.2.[226]



Scheme 2.1- An outline of the research discussed within this chapter; Section 2.1 details the computational design processes including the alignment of the two protein sequences, Section 2.1.6 the three synthetic approaches of the first generation of inhibitors (Section 2.2) and Section 2.3 describes the biophysical techniques used including protein assignment and ligand binding studies by NMR, ITC and fluorescence spectroscopy.

2.1 Computational Design

2.1.1 Introduction

Structure-based drug design is an approach in which computational modelling is used in conjunction with “hit identification” and “hit to lead” optimisation processes in a drug discovery pipeline.[80,81] It is an iterative process that has been developed substantially over the past 40 years and involves progression over multiple cycles incorporating measured experimental data as feedback to optimise future generations of compounds with improved properties.

Once the target has been identified, it is necessary to have structural information. This will either be of the whole protein or the relevant part. The structure can be determined using X-ray crystallography, NMR spectroscopy or homology modelling, as part of the drug design process, or it is available from the Protein Database (PDB).[227,228] Through structure determination with the natural binding partner for example, the residues within the binding site may be targeted in subsequent hit identification steps.[82]

Concerning hit identification for protein-protein interactions, two strategies can be employed. In the first *de novo* design strategy, the lead compounds are built “from scratch” by a computational algorithm in order to identify compounds which are highly complementary to the targeted binding site. This process relies upon the identification and use of the pharmacophore for the target.[229] The second which is virtual high through-put screening, is a computational adaptation of experimental high through-put screening and it involves computationally screening libraries of compounds to fit the requirements of the biological target.[230]

The strategy used in this project is a mixture of both design and virtual high through-put screening; it was initially developed to target the protein-protein interaction that occurs between frq1, the yeast homologue of the neuronal calcium sensor 1 (NCS1) and a peptide from the pik1 protein (Figure 2.1).

As previously discussed in the introduction (Chapter 1.2), a challenge encountered when targeting two interacting proteins with small molecules is that the interface between the two has a tendency to be large and flat.[231] However, there are normally a number of key residues involved in binding one protein to another and these residues are known as “hotspots”. These hotspot residues can be used to develop a pharmacophore and a process known as pharmacophore-based virtual screening can be used to search compound libraries. A pharmacophore was defined as a molecular framework, composed of certain structural elements necessary for its

biological activity.[83,89,232] In virtual screening, a pharmacophore can be used to extract the common features of a set of active molecules and this is known as ligand-based pharmacophore modelling.[83] Structure-based pharmacophore modelling is when the hotspot residues are known and the pharmacophore is used to represent certain points in 3-D space; it could also incorporate certain functional groups to be mimicked. Depending on the type of interaction, these could include hydrogen bond donors, hydrogen bond acceptors, aromatic rings, and charged moieties.[90,91]

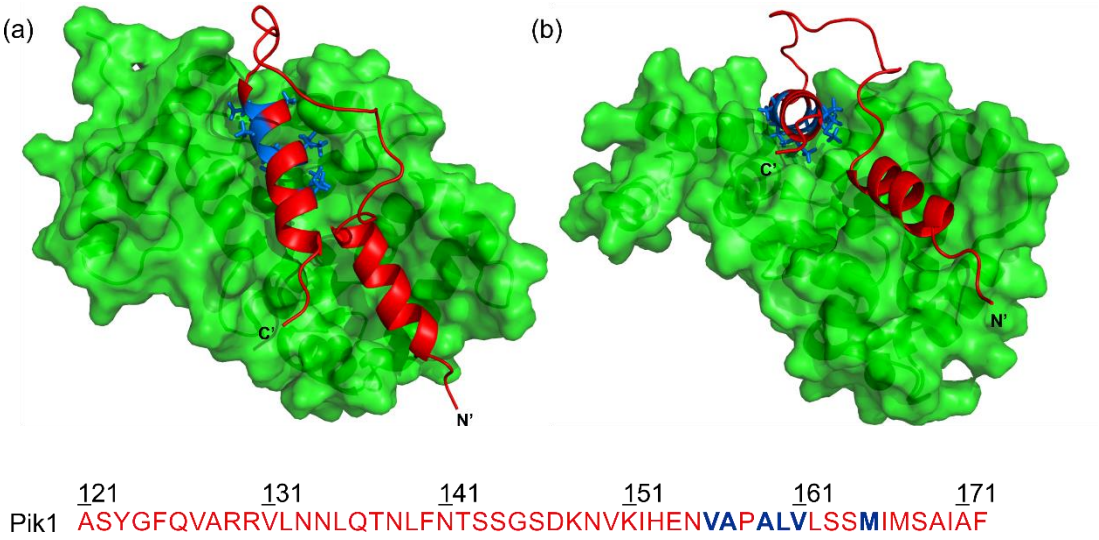


Figure 2.1- The frq1: pik1 complex. Representation of the crystal structure of the frq1: pik1 complex, the model system (PDB 2JU0). Pik1 (red) was shown to anchor itself to frq1 (green) at the C-terminal helix, this hotspot binding interaction involves six key residues on pik1 (coloured blue with the sidechains visible) Val156, Ala 157, Ala159, Leu160, Val161 and Met165.[233] The *C*-terminal helix of pik1 interacts with the *N*-terminus of frq1 and the *N*-terminus of pik1 with the *C*-terminus of frq1. **(a)** View above the hydrophobic. **(b)** Cross section view, the projection of the key amino acid sidechains of the C-terminal helix down into the hydrophobic cleft of frq1. The amino acid sequence of pik1 can be seen below the crystal structures and is coloured red with the 6 key residues coloured in blue.

With the structure of the pharmacophore defined, the virtual high throughput stage can be used to screen a 3-D database for compounds which fit the pharmacophore. There are a number of databases/ compound libraries that can be used. ZINCPharmer is the online interface to one such database ZINC and this online resource conducts pharmacophore searches of a library of purchasable compounds.[92] The interface allows for different filters to be applied to a pharmacophore query; such filters include molecular weight (mw), root mean squared deviation (RMSD) and the number of rotatable bonds, all of which have been implicated in the design process of drug-like molecules.[95]

Searching for compounds that fit a particular pharmacophore is achieved through the assessment of a compounds binding pose with respective protein. This process is performed through docking calculations, and is central to the structure-based design strategy.

There are a number of algorithms that are available to rationalise the interactions between a ligand and a protein.[96] A discussion of the different algorithms can be found in the introduction (Chapter 1.4.3), however the docking programme used for this project was Genetic Optimisation for Ligand Docking (GOLD). Most docking programmes make the approximation that the protein a rigid entity and the ligand as being the flexible component. GOLD, uses a genetic algorithm to investigate the different conformations, placement and rotations of a flexible ligand and takes into account the loss of water from a surface upon ligand binding.[97,109]

As with most docking programmes GOLD is composed some main constituents; 1) a scoring function that is able to rank different binding modes and 2) a mechanism by which it explores the possible binding poses of a ligand.[99] The different docking conformations of each ligand can be assessed for predicted strength of binding using scoring functions such as Astex Statistical Potential (ASP), Chemscore and Chem Piecewise Linear Potential (PLP).[99] ASP uses a protein-ligand database to measure the atom-atom potential, Chemscore estimates the change in energy that occurs upon ligand binding and ChemPLP is an empirical fitness function which models the steric complimentary between the protein and the ligand.[99,111,112]

The docking predictions between a protein and ligand are important in identifying the possible binding interactions. However another vital part of structure-based drug design, is the calculation of a compounds drug-like ADMET properties. The ADMET properties of a compound play a significant role in the overall potential for a molecule to have desired therapeutic potential. Optimisation of the interactions between a ligand and macromolecule target such as a protein, call for alterations of the properties of a small molecule.[113] Historically, there are a number of guidelines for the “ideal drug-like profile” that a compound should possess, this has changed numerous times as knowledge and understanding of macromolecule and ligand interactions has evolved. However despite these continuous changes certain characteristics have remained, for a compound to be successful as a drug it must be well absorbed, metabolically stable, have minimal toxic effects and a good duration of activity.[113]

In order to assess this, computational chemists must take into account a number of physical properties of a compound before it is selected, these include; size, molecular surface area, lipophilic profile (clogP), charge and hydrogen bond donors or acceptors.[113,234] There are a number of chemoinformatic tools that can be used to calculate these parameters; one such programme is Pipeline Pilot which can be

used to generate a sequential workflow able to calculate the required chemoinformatic data (Section 2.1.4).[125]

Overall structure-based drug design is vital in the development of new therapeutics, the computational processes remove the costly high-throughput chemical screening in experimental settings, thereby allowing for a more refined selection of compounds, with a greater chance of becoming a hit to lead success.

2.1.2 Pharmacophore Generation

In designing inhibitors for protein-protein interactions (PPIs), the complex nature of the systems involved means that conventional high through-put screening methods may not be as efficacious in generating hit compounds.[79] With this in mind a computational pipeline was designed utilising a combination of online resources and local modelling algorithms to optimise the chemical structure of the protein with a number of compound. Application of this approach to the model system of frq1 pik1, and informed by Ames *et al.* the identification of a key pharmacophore of pik1 was conducted in PyMOL.[41,224] The pharmacophore was composed of the six hydrophobic residues Val156, Ala157, Ala159, Leu160, Val161, and Met165 (Figure 2.2), the same residues as were indicated by Ames as being important in the binding of pik1 to frq1.[41]

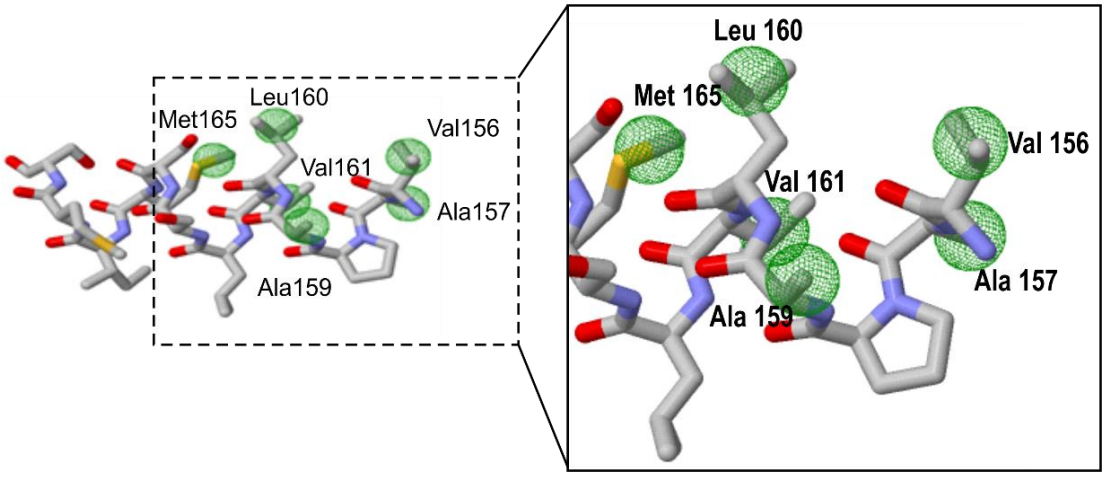


Figure 2.2- A stick representation of a backbone fragment from the *C*-terminal helix of pik1 (PDB 2JU0). The six key amino acid residues of the six point pharmacophore used to screen for compounds in ZINCPharmer labelled. The hydrophobic pharmacophore points are indicated by green spheres.

The pharmacophore screening approach searches for compounds that are able to match closely the desired pharmacophoric array *via* an energetically accessible molecular conformation (Figure 2.2). It was used as an approach to conduct prior to molecular docking, because it was expected that the compounds able to replicate the pharmacophore would have better predicted “strength of binding”, than those randomly selected, importantly pharmacophore searching can be quicker than

docking. The pharmacophore approach is an informed, data driven approach to identify compounds that could bind as desired. Those compounds once identified undergo more thorough assessment *via* docking processes.

Combinations of residues were varied systematically (as described below), and the selection decided by manual investigation and visualisation of the structure, which was further informed by what is known in the literature.[41] This resulted in six different pharmacophore's (Table 2.1) that were used as queries in ZINCPharmer.[224] By selecting various combinations of residues involved in the pharmacophore definition, the size, shape (brought about through different substitution patterns) and chemical functionality of the compounds identified from each pharmacophore search will be different from one another. For example inclusion of all 6 residues in a pharmacophore search may result in hit compounds that are slightly larger than those where the pharmacophore involves just four consecutive residues. The minimum number of residues included in the pharmacophore selection was four, as it was suggested that compounds that only fit a three point pharmacophore or less may be too small, and result in a non-specific interaction in the resultant testing.[235] The composition of the residues involved in the six different pharmacophore's for each enquiry can be found in Table 2.1.

Table 2.1- **Pharmacophore generation**. A table highlighting the combination of the six key residues from the pik1 fragment that made up each enquiry for the pharmacophore searches. The residues included in each enquiry are highlighted in red.

Residue:	Val156	Ala157	Ala159	Leu160	Val161	Met165
Enquiry Number						
1						
2						
3						
4						
5						
6						

The ZINCPharmer database contains ≃ 35 million ligands and ≃ 200 million ligand energetically accessible conformations and is updated regularly.[224] The molecular weights filter of the results was set such that they never exceeded 600 Da, as larger compounds may present challenging pharmacokinetic profiles.[43,44] The root mean squared deviation (RMSD) is a value based upon the difference between pharmacophoric points in the query and those presented by a particular molecule's conformation, therefore smaller RMSD values are preferred and a maximal threshold value of 2Å was used.[236] The number of rotatable bonds is a measure of the molecular flexibility of a compound; it has been shown to be an important indicator of

oral bioavailability and includes any single bonds not bonded to a hydrogen atom, not involved in a ring system and does not include amide bonds.[95]

The parameters applied to each pharmacophore search were as follows: molecular weight (Da) ≤ 600, ≤ 500, ≤ 300 and ≤ 150; RMSD ≤ 2, ≤ 1 and ≤ 0.5 and rotatable bonds ≤ 9 and between 9 and 1. The results of the ZINCPharmer searches can be seen tables 2.2 - 2.7, the numbers within the table under each heading represent the number of molecules for that individual enquiry. For example; enquiry 1 (Table 2.2) where the pharmacophore included the residues Val156, Ala159, Leu160 and Met165, those compounds within filter (a) with a molecular weight ≤600 Da, RMSD ≤2 and number of rotatable bonds ≤9 included 1,302,105 compounds. As the molecular weight, RMSD and number of rotatable bond values were decreased (filters b-k Table 2.2) so did the number of hit compounds.

Table 2.2- **Enquiry 1 results.** The table showing the ZINCPharmer results for Enquiry 1 targeting the residues: Val156, Ala159, Leu160 and Met165. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 1,302,105 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filter	Rotable bonds	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5
a	≤9	1,302,105	1,302,105	66,242	1,244,140	1,244,140	63,626	13,127	13,127	404	0	0	0
b	9	311,078	311,078	17,715	292,519	292,519	16,642	2,527	2,527	92	0	0	0
c	8	306,982	306,982	16,442	291,647	291,647	15,604	2,935	2,935	109	0	0	0
d	7	262,413	262,413	13,328	251,211	251,211	12,808	2,900	2,900	81	0	0	0
e	6	198,128	198,128	9,398	191,404	191,404	9,097	2,057	2,057	63	0	0	0
f	5	125,457	125,457	5,593	121,900	121,900	5,470	1,461	1,461	39	0	0	0
g	4	64,836	64,836	2643	63,075	63,075	2,546	854	854	16	0	0	0
h	3	26,301	26,301	916	25,542	25,542	901	297	297	2	0	0	0
i	2	6,207	6,207	187	6,096	6,096	181	86	86	2	0	0	0
j	1	646	646	18	635	635	17	10	10	0	0	0	0
k	0	57	57	2	51	51	2	0	0	0	0	0	0

Table 2.3- **Enquiry 2 results.** The table showing the ZINCPharmer results for Enquiry 2 targeting the residues: Val156, Ala157, Ala159 and Leu160. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 3,162,529 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filter	Rotable bonds	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5	≤ 2	≤ 1	≤ 0.5
a	≤9	3,162,529	3,162,529	210,125	2,919,509	2,919,509	192,082	22,395	22,395	849	0	0	0
b	9	934,000	934,000	62,110	852,215	852,215	56,203	5,737	5,737	249	0	0	0
c	8	839,883	839,883	56,760	773,121	773,121	52,718	6,199	6,199	250	0	0	0
d	7	665,154	665,154	44,552	619,360	619,360	41,240	5,373	5,373	201	0	0	0
e	6	421,407	421,407	27,712	389,217	389,217	25,397	3033	3033	83	0	0	0
f	5	206,455	206,455	12,998	192,054	192,054	11,962	1360	1360	47	0	0	0
g	4	73,255	73,255	4,451	67,628	67,628	4,102	572	572	14	0	0	0
h	3	20,500	20,500	1,216	19,450	19,450	1,151	97	97	4	0	0	0
i	2	6,411	6,411	305	6,080	6,080	290	23	23	1	0	0	0
j	1	318	318	20	296	296	19	1	1	0	0	0	0
k	0	146	146	1	88	88	88	0	0	0	0	0	0

Table 2.4- **Enquiry 3 results.** The table showing the ZINCPharmer results for Enquiry 3 targeting the residues: Ala157, Ala159, Leu160 and Val161. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 2,581,274 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filter	Rotable bonds	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5
a	≤9	2,581,274	2,581,274	172,678	2,392,764	2,392,764	159,116	17,422	17,422	563	0	0	0
b	9	765,172	765,172	49,682	102,373	102,373	45,256	4,447	4,447	159	0	0	0
c	8	689,234	689,234	146,324	638,443	638,443	42,426	4,790	4,790	188	0	0	0
d	7	536,698	536,698	36,791	50,403	50,403	34,352	4,167	4,167	138	0	0	0
e	6	335,574	335,574	22,968	312,269	312,269	21,209	2,297	2,297	48	0	0	0
f	5	168,251	168,251	11,329	156,806	156,806	10,635	1,043	1,043	21	0	0	0
g	4	61,775	61,775	4,181	57,331	57,331	3,904	521	521	9	0	0	0
h	3	18,328	18,328	1,068	17,287	17,287	1,018	122	122	0	0	0	0
i	2	5,821	5,821	316	5,515	5,515	303	33	33	0	0	0	0
j	1	284	284	14	260	260	13	2	2	0	0	0	0
k	0	137	137	5	77	77	0	0	0	0	0	0	0

Table 2.5- **Enquiry 4 results.** The table showing the ZINCPharmer results for Enquiry 4 targeting the residues: Val156, Ala157, Ala159, Leu160 and Met165. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 12,351 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filters	Rotable bonds	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5
a	≤9	12,351	12,351	191	10,694	10,694	157	14	14	0	0	0	0
b	9	4,691	4,691	58	4,119	4,119	44	4	4	0	0	0	0
c	8	3,681	3,681	55	3,229	3,229	47	6	6	0	0	0	0
d	7	2,275	2,275	41	1,929	1,929	39	4	4	0	0	0	0
e	6	1,228	1,228	21	1,024	1,024	13	0	0	0	0	0	0
f	5	333	333	8	270	270	7	0	0	0	0	0	0
g	4	105	105	6	87	87	5	0	0	0	0	0	0
h	3	31	31	0	30	30	0	0	0	0	0	0	0
i	2	7	7	2	6	6	2	0	0	0	0	0	0
j	1	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.6- **Enquiry 5 results.** The table showing the ZINCPharmer results for Enquiry 5 targeting the residues: Val156, Ala159, Leu160 and Val161. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 3,609 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filter	Rotable bonds	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5
a	≤9	3,609	3,609	88	3,215	3,215	84	0	0	0	0	0	0
b	9	1289	1289	44	1,132	1,132	41	0	0	0	0	0	0
c	8	954	954	14	850	850	14	0	0	0	0	0	0
d	7	763	763	27	687	687	27	0	0	0	0	0	0
e	6	305	305	3	261	261	2	0	0	0	0	0	0
f	5	241	241	0	232	232	0	0	0	0	0	0	0
g	4	34	34	0	30	30	0	0	0	0	0	0	0
h	3	22	22	0	22	22	0	0	0	0	0	0	0
i	2	1	1	0	1	1	0	0	0	0	0	0	0
j	1	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 2.7- **Enquiry 6 results.** The table showing the ZINCPharmer results for Enquiry 6 targeting the residues: Val156, Ala159, Leu160, Val161 and Met165. This table shows the number of compounds under each search criteria, for example Filter a- compounds with a molecular weight ≤600 Da, RMSD ≤2 Å and number of rotatable bonds ≤9 gives 21 compounds.

		Molecular Weight ≤ 600			Molecular Weight ≤ 500			Molecular Weight ≤ 300			Molecular Weight ≤ 150		
		RMSD			RMSD			RMSD			RMSD		
Filter	Rotable bonds	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5	≤2	≤1	≤0.5
a	≤9	21	21	0	18	18	0	0	0	0	0	0	0
b	9	8	8	0	5	5	0	0	0	0	0	0	0
c	8	11	11	0	11	11	0	0	0	0	0	0	0
d	7	2	2	0	2	2	0	0	0	0	0	0	0
e	6	0	0	0	0	0	0	0	0	0	0	0	0
f	5	0	0	0	0	0	0	0	0	0	0	0	0
g	4	0	0	0	0	0	0	0	0	0	0	0	0
h	3	0	0	0	0	0	0	0	0	0	0	0	0
i	2	0	0	0	0	0	0	0	0	0	0	0	0
j	1	0	0	0	0	0	0	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0	0	0	0	0	0

The results indicate that the number of hits did not alter between the searches of max RMSD ≤ 2 Å and ≤ 1 Å in all enquiries, which suggested that the optimal RMSD range was be between 1 and 0.5 Å as the values between the two search parameters differed. The number of hits was greatly reduced after limiting the search to compounds with a molecular weight ≤ 300 Da, suggesting that the key molecular weight range for maximising potential diversity of compounds discovered was between 600 - 300 Da, as a small compound would not be able to reach the pharmacophoric points which are far from each other. This result is in line with previous studies which suggested that PPI inhibitors do not follow Lipinski's "rule of 5" and generally had molecular weights greater than 500 Da which could be explained due to the large size and topography of the binding site between two interacting proteins.[49,231] The sixth enquiry saw a dramatically reduced number of hits with the broadest parameter setting of mw ≤ 600 Da, RMSD ≤ 2 Å and rotatable bonds ≤ 9. This suggests that the larger the number of pharmacophore points (amino acid residue targets) the fewer the number of possible compounds that are a good match; this could be useful for specificity. The specificity of a compound is important in minimising the chance of off-target interactions.[237]

In order for the ranking and selection of the most suitable compounds, a balanced selection process using several inputs (predicted strength of binding and predicted ADMET properties) was applied in order to select the most viable ligands for consideration for synthesis (Figure 2.3). It follows an iterative process and involves, docking and scoring the compounds, calculation of the ADMET profiles to allow for the generation of the compound with the most suitable binding and "drug like" profile.

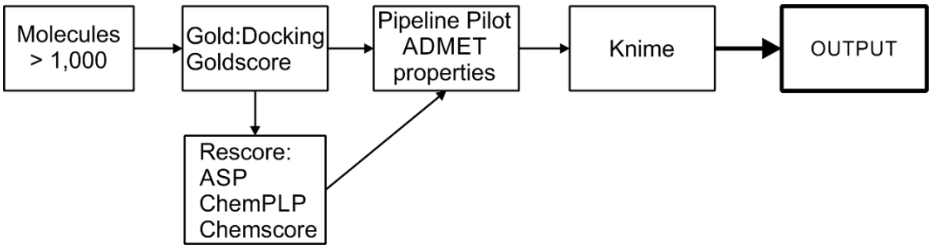


Figure 2.3- The overall approach of the balanced selection workflow for the larger enquiries. Which involved taking the molecules from the online pharmacophore screening process, docking the compounds into frq1 using the docking software GOLD and utilising the scoring function Goldscore. The Goldscore docking poses of each compound were subsequently re-scored in GOLD using the scoring functions ASP, ChemPLP and Chemscore (each scoring function works on different parameters see GOLD Docking Section 2.1.3 for more information). The compounds ADMET properties were calculated, this along with each of the four scoring values were ranked in Knime and the ligand efficiency values calculated. This was conducted in order to select the compounds with the best overall chemical profile and indicate which should be visually inspected.

2.1.3 GOLD Docking Analysis

Following the identification of compounds which matched various pharmacophores, the next step in the computational workflow was docking and scoring each of the compounds contained in tables 2.2 – 2.7 a- k (above) into frq1 using GOLD.[99]

One particular enquiry, enquiry 6 (Table 2.7) contained only 21 virtual hit compounds from the Zinc searches; this enquiry was manually assessed and so will be discussed separately (Section 2.1.5).

The process of docking involves the prediction of the conformation and orientation of ligand's within the binding site.[238] Initially the ligands in each file were scored using the Goldscore which generated a fitness value based upon the protein-ligand hydrogen bonding energy, protein-ligand Van Der Waals (VDW) energy, the ligand internal VDW energy and the internal torsional strain of the ligand.[99] The better Goldscore's are those of a higher value; they can subsequently be used to generate the ligand efficiency of each compound.

Ligand efficiency is an estimation of the amount of binding per atom of the ligand with the target protein; an example of the calculated Goldscore and ligand efficiency values for the compounds within enquiry 6 can be seen in Table 2.8.[239] As medicinal chemists strive to increase the potency of lead compounds, the optimisation processes involved in drug development often result in increasing molecular weight.[240] The ligand efficiency is a simple metric that is important to optimise potency, whilst maintaining optimal molecular size for a "drug-like" compound.[239]

The predicted Goldscore conformations of each compound were consequently rescored in GOLD using three different scoring functions, ASP, Chemscore and ChemPLP. The consensus scoring process was used to ascertain an overall

assessment of the interaction profile between each compound and the target protein.[99]

Consensus scoring is used to improve the hit rates when screening a library of compounds; it is a simple rescoring and re-ranking of docking solutions based on a different scoring function.[99] The consensus scoring method is statistically improved, the mean of the samples is closer to a "true" sample than an individual value.[241] The different methods focus on different aspects of ligand binding and so agree on an active compound more than an inactive one as well as providing a greater consistency across receptor systems.[241] The consensus scoring approach has been indicated as leading to improvements in docking procedures, through the improved quality of results with less reliance on the correct selection of a scoring function.[242] As such this consensus approach to the docking process of the ligands allows for a more informed selection that is not biased, but rather incorporates as many possible aspects of a binding interaction as possible.

A subsequent process in the computational analysis was the docking of pik1 into frq1 using GOLD to calculate its ligand efficiency from its Goldscore so to generate a benchmark for the compounds of each enquiries. The key residues of frq1 identified by Ames were highlighted in a PyMOL representation of the protein (Figure 2.4 a) [224] and the top docking pose of the pik1 fragment was then read into the file to observe the key interactions between the protein and peptide (Figure 2.4 b).

The drug-like profile of each compound is also necessary to predict their oral bioavailability and toxicity, to do this the solubility, surface area and volume, number of hydrogen bond donors and acceptors, ClogP and molecular weight were calculated. The link between ADMET and the pharmacokinetic descriptors has been investigated extensively, one such investigation at GSK described how the variations of 12 molecular descriptors including hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), molecular weight (MW), polar surface area (PSA) and rotatable bonds (bonds) could be described with four orthogonal components.[118] The group found that of the four orthogonal components, molecular weight and ClogP were the two main characteristics that determine ADMET liabilities.[118] The investigation detailed how the changes in the two key molecular properties, will affect a range of ADMET parameters including solubility, bioavailability permeability, volume of distribution, P450 inhibition (relating to metabolism) and *in vivo* clearance.[118]

Table 2.8- **Enquiry 6 docking results.** The Goldscore and ligand efficiency values of the compounds from enquiry six (Table 2.7), the data set that was used to select the inhibitors for synthesis and testing, the ligand efficiency values are calculated as average Goldscore / molecular weight.

Zinc code	Average Goldscore	Ligand efficiency
Pik1		0.0375
ZINC09276881	39.46	0.08
ZINC09276880	41.23	0.08
ZINC04792655	35.45	0.07
ZINC01068766	40.45	0.08
ZINC03191471	38.63	0.10
ZINC03191471	37.89	0.10
ZINC65321569	39.71	0.09
ZINC12140543	41.37	0.08
ZINC16268178	39.53	0.08
ZINC24198265	42.88	0.09
ZINC09276852	39.91	0.09
ZINC09276853	40.57	0.09
ZINC08812681	39.94	0.09
ZINC09668287	39.33	0.09
ZINC03403995	39.67	0.09
ZINC20351159	37.92	0.08
ZINC12242584	35.51	0.08
ZINC68119468	37.56	0.10
ZINC66781206	39.87	0.10

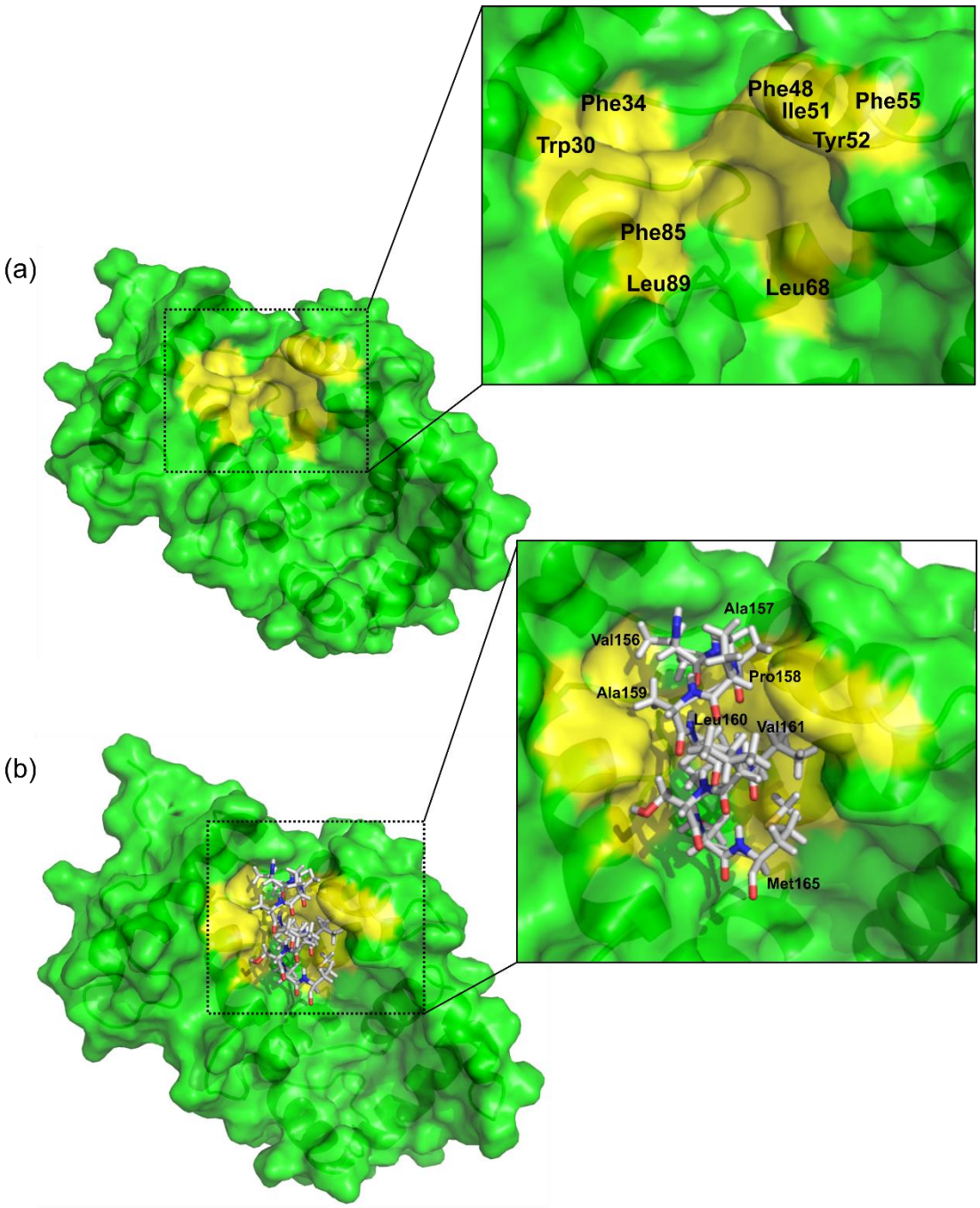


Figure 2.4- **The GOLD docking poses of the frq1/ pik1 complex.** (PDB 2JU0) **(a)** Frq1 (green) with the key hydrophobic residues (yellow) indicated by Ames *et al.* [4] as being the key hotspot of the protein for its interaction with its binding partner pik1 these residues include Trp30, Phe34, Phe48, Ile51, Tyr52, Phe55, Leu68, Phe85 and Leu89. **(b)** The protein-protein interaction between frq1 and the six hydrophobic residues Val156, Ala157, Ala159, Leu160, Val161, and Met165 from the *C*-helix of pik1 thought to be necessary for anchoring the two proteins.

2.1.4 Physicochemical property determination and balanced selection

Calculation of the physicochemical properties of each of the molecules was conducted in the scientific dataflow platform, Pipeline Pilot.[125] Using Lipinski's parameters as a guideline, a work flow was generated as a strategy to calculate specific physicochemical properties which influence ADMET properties. Individual nodes were used to calculate, molecular weight, Log P, number of hydrogen bond acceptors and donors as well as the number of rotatable bonds, solubility and polar surface area for each compound from the pharmacophore searches (Figure 2.5).

It is beneficial to calculate the physicochemical properties before synthesis of any compounds, as it can be used to predict the likely action of the body on the compound. Should the ligand be intended for use as a therapeutic agent, it may help to deselect those ligands with unfavourable ADMET profiles.[113] An important physicochemical property is the solubility, not only is this important if the compounds are designed as a therapeutic agent, but also for the purposes of any biophysical analysis. A compound with a poorly soluble profile is likely to pose problems in both testing and future rounds of the molecular design process.

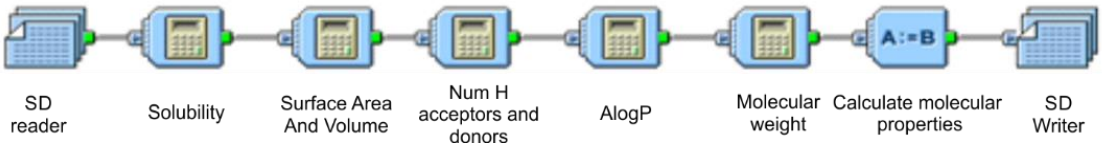


Figure 2.5- The sequential pipeline pilot workflow used to calculate the ADMET properties of each compound. The data generated from this workflow was subsequently used to further filter the compounds and lead to the selection of a number of small molecules to synthesise.

The files were then processed in Knime filtering off any undesirable ligands, those with negative docking fitness scores. These compounds were disregarded as a negative fitness score relates to an extremely unfavourable predicted binding pose/ clash with protein.

An overall average fitness score was calculated for each compound from the four scoring functions generated in GOLD, the ligand efficiencies of each molecule were calculated and then sorted into descending order of this value. The ligand efficiency was then compared to that of the value calculated from the docking of pik1, the natural peptide ligand. Molecules would be carried forward if they possessed better ligand efficiency than pik1 and these compounds were ranked using a *Pareto* optimisation.

Pareto ranking is a method by which a compound is ordered with respects to the number of other compounds that are able to dominate it with respects to a number of objectives. It is a multi-objective optimisation process to determine the overall

characteristics of a compound which can include their overall scoring function for example.[243] The *Pareto* settings applied were as follows, molecular solubility was maximised, an optimal range of the number of hydrogen bond acceptors and donors was set as 1.0 - 10.0 and 1.0 - 5.0 respectively, in keeping with the Lipinski guidelines.[184] The optimal range of A Log P was set to a range of 1.0 – 3.5 in order to maximise the solubility profile. Consensus ranking and molecular weight were selected to be minimal values to improve the chances of a compound with good ADMET attributes. This ranked list of ligands produced a single output file which could be visually inspected and analysed in PyMOL with frequenin as the binding partner (the Knime workflow system can be seen Figure 2.6).[123,224]

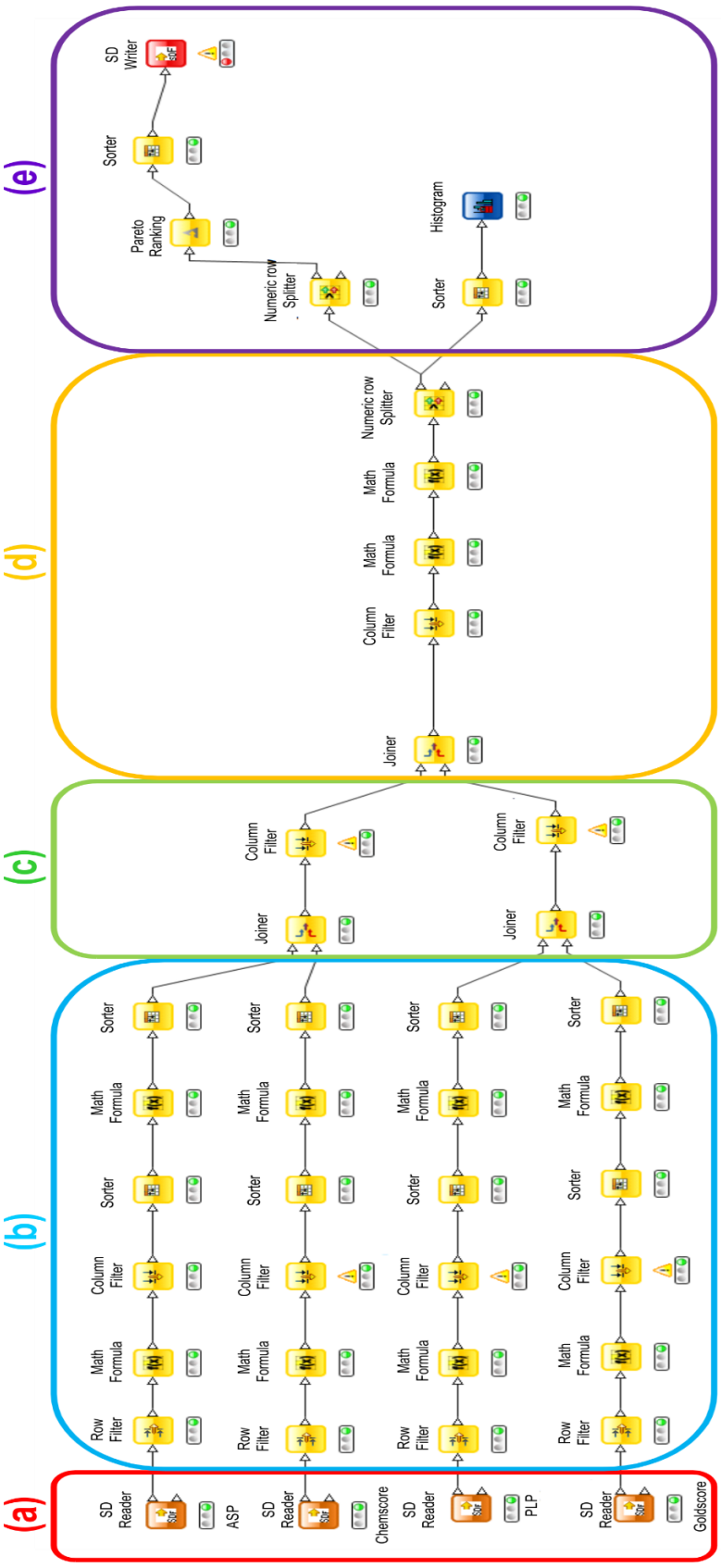


Figure 2.6- The Knime workflow applied to those enquiry's that contained an extremely large number of compounds for which further ranking and filtering was necessary before the selection of compounds to be visually inspected was carried out. **(a)** Each four docking files from GOLD are read into Knime in SDF format and the files contain the calculated Physiochemical properties. **(b)** Any compounds with negative fitness scores are filtered out straight away as they are unfavourable. A math formula is applied to calculate the ligand efficiency of each compound; further filters are applied with outputs containing "Molecule Name", "Fitness Score Solubility", "Surface Area and Volume", "Number of Hydrogen Acceptors and Donors", "A LogP", "Molecular Weight", "number of rotatable bonds" and "Ligand efficiency". The compounds are listed in descending order of Ligand efficiency and a math formula is applied to calculate an overall score of the chemical profile for each ligand, which are then ranked in order with respects to its new score. **(c)** The data from each of the four scoring files is then merged and duplicate data is filtered. **(d)** A "better than pik1" score is calculated through comparison of the ligand efficiency of each compound to that of the native pik1 peptide. **(e)** A *Pareto* ranking system is used to list the compounds with the best overall profiles and this includes the compounds with the best overall scoring values and ligand efficiencies.

2.1.5 Enquiry Six

The pharmacophore of enquiry number six included all six residues of the pik1 fragment: Val156, Ala159, Leu160, Val161 and Met165, indicated by Ames *et al*. as being key to anchoring the *C*-terminal helix of pik1 to frq1. This topologically large pharmacophore significantly restricted the number of compounds that were able to successfully match the pharmacophoric points.

Twenty one virtual hits were identified from the pharmacophore search, although on closer inspection two compounds were conformers of the same compound. The small number of molecules within enquiry six meant that selection of possible compounds was carried out manually without the use of Knime. It involved analysis of each compound's scoring function, along with their physicochemical properties and commercial availability. Physicochemical analysis of enquiry 6 and the commercial availability of the compounds filtered out 11 of the 19 ligands. Removing those with unfavourable molecular weights of around 600 Da and those which appeared commercially unavailable, resulted in 8 possible hit compounds. The top scoring binding poses of the eight compounds were then loaded into the binding pocket of frq1 in PyMOL, to see how they were predicted to interact with frq1 and were compared to the binding orientation of pik1 around the key residues (Figure 2.8). The top binding pose of ligands, were then overlaid with the pik1 pharmacophore and through visual analysis their relative merits were assessed (Table 2.9).

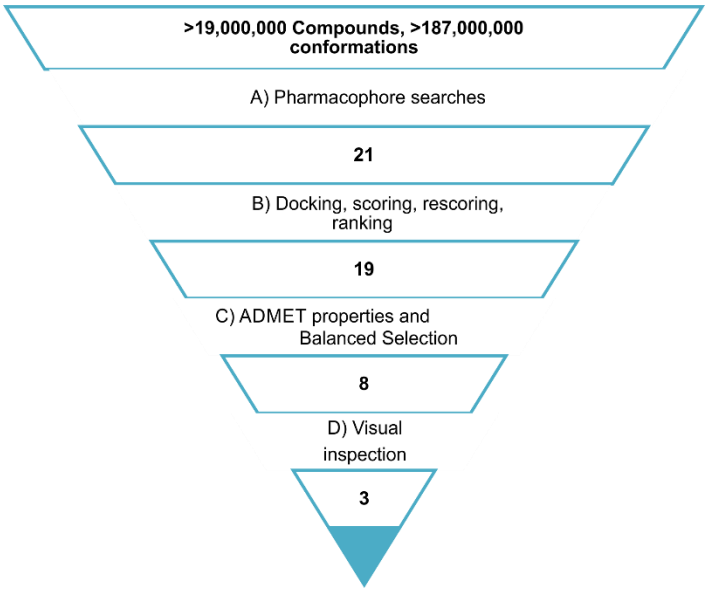


Figure 2.7- The computational work flow of Enquiry 6. Initial screening of the millions of compounds from the online database with the pharmacophore Val156, Ala159, Leu160, Val161 and Met165 led to 21 hit compounds. Subsequent docking, scoring and other computational screening strategies resulted in eight target compounds that were analysed visually through inspection of pharmacophore target and the small molecules. This resulted in the selection of three compounds from the original 21 for synthesis and further biophysical analysis (to be discussed in later sections 2.2 and 2.3)

Table 2.9- Visual analysis of the six key residues of pik1. A comparison of the 3-D space occupied by the compounds. The pharmacophore of enquiry 6 (table 2.7) overlaid with the eight compounds, those residues whose area in 3-D space appeared to be mimicked by a region of a specific compound are highlighted in red.

Residue:	Val156	Ala157	Ala159	Leu160	Val161	Met165
Compound **2.1**						
Compound **2.2**						
Compound **2.3**						
Compound **2.4**						
Compound **2.5**						
Compound **2.6**						
Compound **2.7**						
Compound **2.8**						

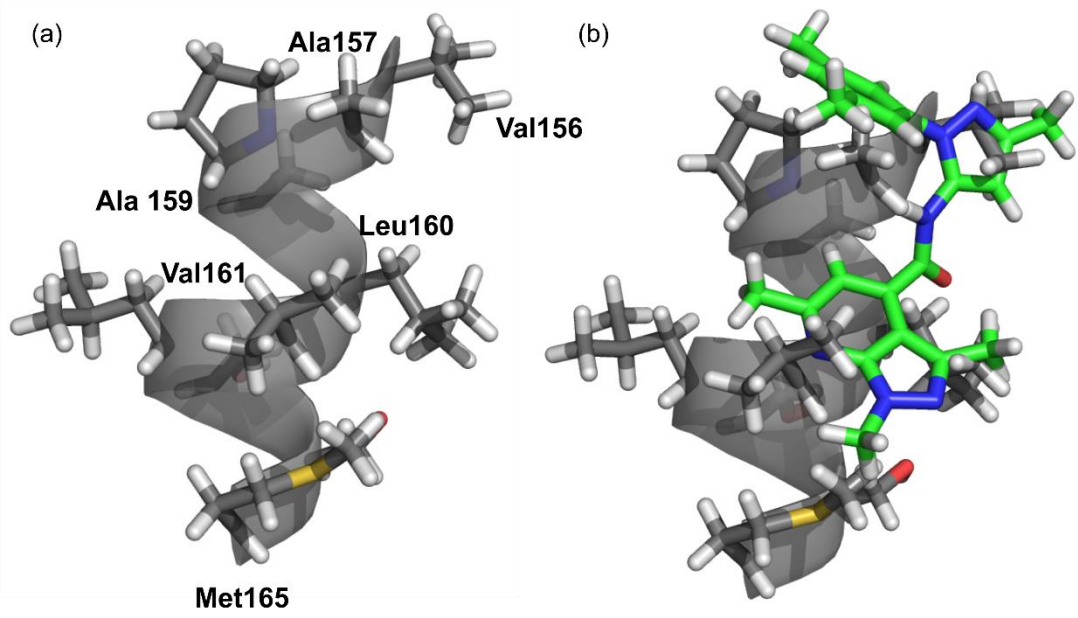


Figure 2.8– Overlays between the compound structures and the pik1 peptide from the frq1 complex. (a) The pik1 pharmacophore used for the enquiry 6 selection with the six key residues labelled. **(b)** An overlay of the pik1 pharmacophore with compound **2.7**; Met165 and Ala159 are unsuccessfully mimicked; for Met165, this is possibly due to the fact that it is four residues away from the other key residues.

These results show clearly Met165 has not been successfully mimicked by any of the docking solutions of the ligands from Enquiry 6 which is not surprising, as it is four residues away from the other residues. However compounds **2.5** – **2.8** mimicked the greatest number of residues overall. This filtered out compounds **2.1** - **2.4** as they only hit one of the target residues each. Purchasing the compounds was not possible within the financial remits of this research and so the remaining four were analysed retro-synthetically, in order to find a plausible synthetic route to each one. It was most difficult to design synthetic routes for Compounds **2.5** and **2.6** (e and f Figure 2.9). Whereas it was possible to find synthetic routes for Compounds **2.7** and **2.8** (g and h Figure 2.9).

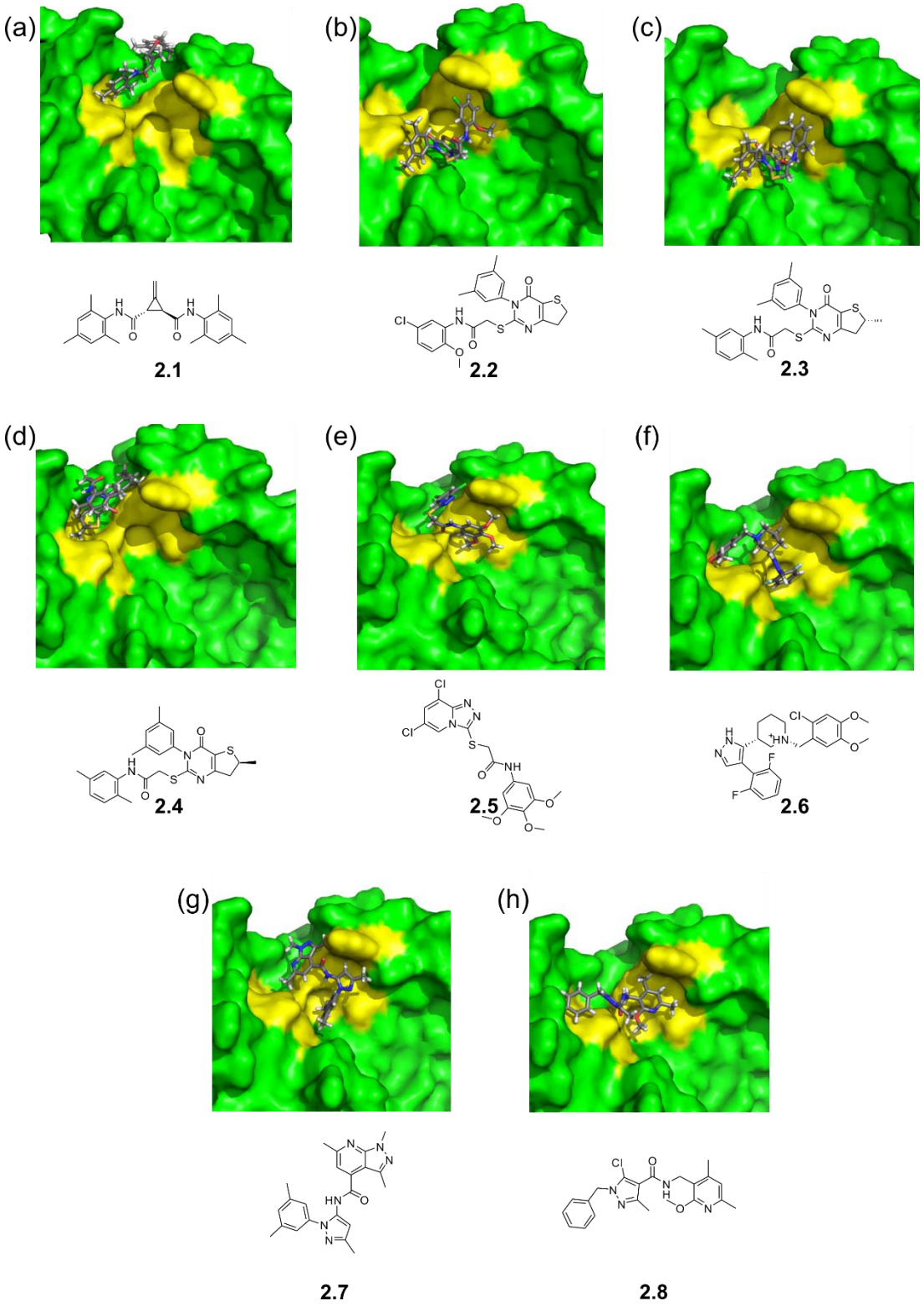


Figure 2.9- The top docking poses of the first generation compounds 1 - 8 from Enquiry 6 generated using the docking programme GOLD. Frq1 can be seen coloured in green with the key hydrophobic residues thought to anchor pik1 to frq1 coloured in yellow.

2.1.6 Applications of a frq1 inhibitor for rat NCS1

At the initiation of this project the crystal structure of the NCS1 D2R peptide complex was unavailable; hence, the structure of the yeast orthologue frq1 in complex with the pik1 peptide (PDB 2JU0) was used to develop the strategy of virtual ligand screening as described above. Both frq1 and NCS1 are predominantly α helical in nature, containing around 11 α helices, 4 β sheets which along with loop regions make up the characteristic 4 EF hand motifs.[38] To ensure frq1 was suitable to use as a model system a comparison between the two sequences of NCS1 and frq1 was carried out comparing and contrasting the binding interactions and determine a possible conserved binding motif.

The alignment of frq1 and NCS1 found that the pair shared a 60% sequence identity and an overall sequence similarity of 75%. Of the eight hotspot amino acid residues between the interactions of the *C*-terminal helix of pik1 with the *N*-terminal helix of frq1 (Figure 2.10), seven are conserved in rat NCS1 and Leu68 is substituted for a valine.

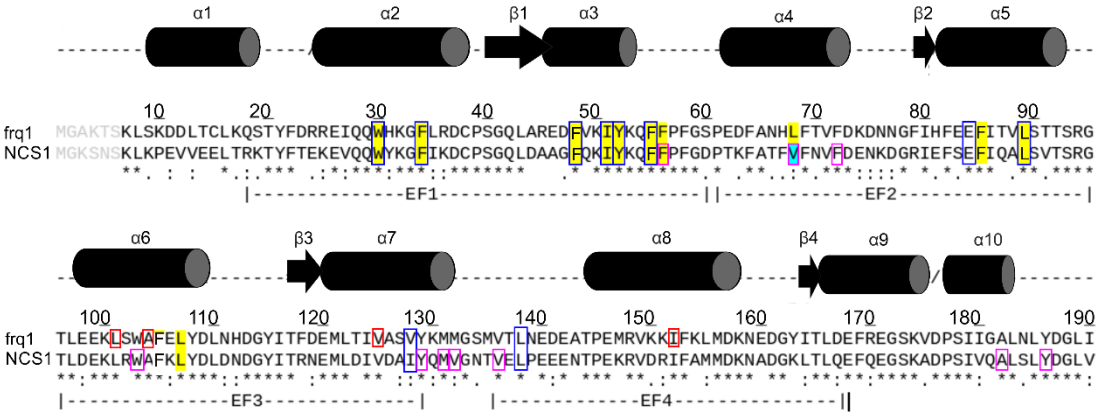


Figure 2.10 -Sequence alignment of Rat NCS1 and Yeast frq1. Aligned sequences of frq1 (PDB 2JU0) and NCS1 (PDB 5AER) with an average combined secondary structure, cylinders represent α helices and arrows β sheet. Overall alignment of the two proteins is indicated as follows; * represent conservation of a residue **":"** and **"."** indicate that the residues are of a similar nature. The residues coloured in grey are the cloning artefacts, the top lines are the frq1 sequence from *Saccromyces cerevisiae*, and those residues highlighted in yellow represent the key hydrophobic residues of frq1 that are important for the binding interaction with pik1. The second sequence is that of NCS1 from *Rattus Norvegicus* used in our experiments. Both human and *Rattus Norvegicus* NCS1 share a 100% sequence identity. The key residues of frq1 found to anchor the C-terminal helix of pik1 were transposed onto NCS1, those in yellow indicate a conservation of sequence and cyan indicates the presence of a different amino acid. The positions of the four EF hand motifs within the structure are also displayed. Those residues in of frq1 that make contacts with pik1 are found within a red box, those of NCS1 that make contacts with the D2R peptide within the magenta boxes, finally the conserved residues of both proteins that make contacts with the respective binding partners within the blue box.

Figure 2.11- Crystal structure of NCS1 highlighting the changes in residues between NCS1 and frq1. NCS1 PDB 5AER with those residues of NCS1 conserved with the residues of frq1 coloured in red, those residues that are of a similar nature coloured in yellow and those residues that are completely different coloured in cyan. The sequence identity between NCS1 and frq1 is 60% and the overall similarity stands at 65%.

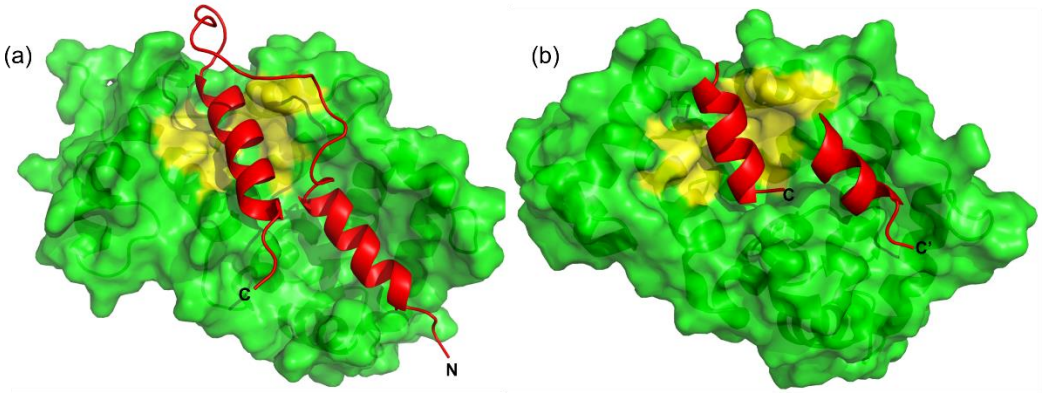


Figure 2.12- Comparison between crystal structures of frq1 and NCS1 with their native ligands. (a) The crystal structure of the frq1 pik1 complex downloaded from the pdb, code 2JU0. Frq1 coloured in green with the single pik1 molecule coloured in red and the key residues of frq1 yellow. **(b)** The crystal structure of the NCS1 D2R peptide complex downloaded from the protein data bank (pdb code 5AER), NCS1 can be seen coloured in green and the two C-terminal helices the D2R peptides can be see coloured in red, one is labelled c and the other c' to indicate that they are from two different D2R molecules.

After selecting inhibitors based on the model system of the frq1 pik1 pharmacophore and with the release of our NCS1 D2R peptide complex structure (PDB 5AER),[13] it was essential to ensure these molecules would also target NCS1.

Interestingly when we compare the interactions of the compounds with frq1 to that of its mammalian orthologue, we observe different binding poses, although crucially they occur in the same hydrophobic binding groove. Compounds **2.5**, **2.7** and **2.8** can be seen to interact with the respective key residues on NCS1 but in different orientations (Figure 2.13). With this data in hand, the analysis of the top binding pose of each of the three ligands with NCS1, can be used to target the protein-protein interaction between NCS1 and the dopamine receptor D2R.

For the purposes of the synthesis, compound **2.8** was modified such that the chlorine group on the pyrazole ring (Figure 2.9, h) was replaced with a second methyl group; this was due to the commercial availability of the compounds used in the synthetic approach. The synthesis of the compounds **2.7**, **2.8** and **2.5** can be found in Section 2.2, where they are now referred to, respectively, as Inhibitor 1, Inhibitor 2 and Inhibitor 3.

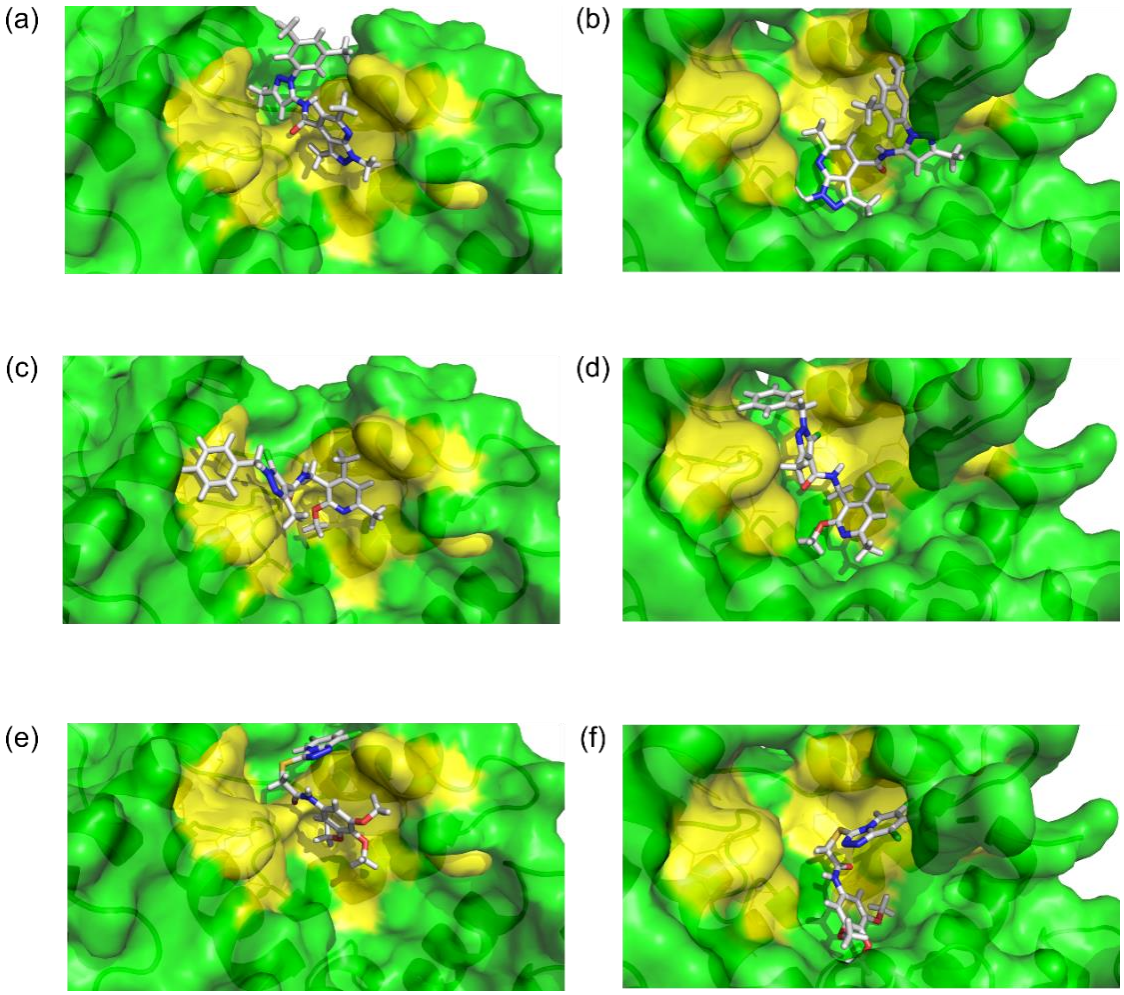


Figure 2.13- Comparisons between the top docking poses of Compounds **2.7**, **2.8** and **2.5** in frq1 PDB 2JU0 (a, c, e) and NCS1 PDB 5AER (b, d, f). The key residues of the frq1 pik1 interaction highlighted in yellow. **(a)** Compound **2.7** top docking pose in frq1 generated in GOLD. **(b)** Compound **2.7** top docking pose in NCS1 generated in GOLD. **(c)** Compound **2.8** top docking pose in frq1 generated in GOLD. **(d)** Compound **2.8** top docking pose in NCS1 generated in GOLD. **(e)** Compound **2.5** top docking pose in frq1 generated in GOLD. **(f)** Compound **2.5** top docking pose in NCS1 generated in GOLD.

2.1.7 Summary of Computational Approach

The computational approach of this project was to identify hit small molecule compounds which would modulate protein-protein interactions. The initial discovery pipeline was developed using the model system frq1 and pik1, this was then transferred to a therapeutic target of the mammalian orthologue of frq1, NCS1 and its binding partner D2R peptide. The rationale is to design a ligand that may bind to all NCS1 orthologues, in order to target conserved and therefore more evolutionary relevant interactions. This interaction is of significant biological relevance as it has been implicated in schizophrenia and bi-polar disorders.[244] The novel pipeline utilised a mixture of structure-based design and high through-put screening. Through combining the two techniques the screening process we would targeting the desired protein efficiently and generating drug-like hit compounds.

With this first generation of inhibitors, the online interface ZINCPharmer allowed a library of purchasable compounds to be screened to fit a specific pharmacophore. The pharmacophore enquiries included a number of screening parameters such as molecular weight, AlogP, number of hydrogen bond donors and acceptors and molecular surface area, all of which have been implicated in improving the solubility profile of drug-like molecules and were used to remove any compounds with undesirable physiochemical parameters.

Pharmacophore virtual screening was used to determine compounds to fit the specific points in 3-D space representing the six key residues of the pik1 fragment (Figure 2.2). Interestingly it was found that by increasing the number of the residues involved in the pharmacophore, the resultant number of hit compounds was significantly reduced (enquiry 6). Enquiries 1 – 5 had resulted in millions of hit compounds and although they did undergo the rigorous docking, physicochemical calculations and *Pareto* ranking of the pipeline, the top compounds were not investigated any further due to time constraints. However, future work could include revisiting these compounds to see if there are any that are predicted as having a better binding profile than the current targets.

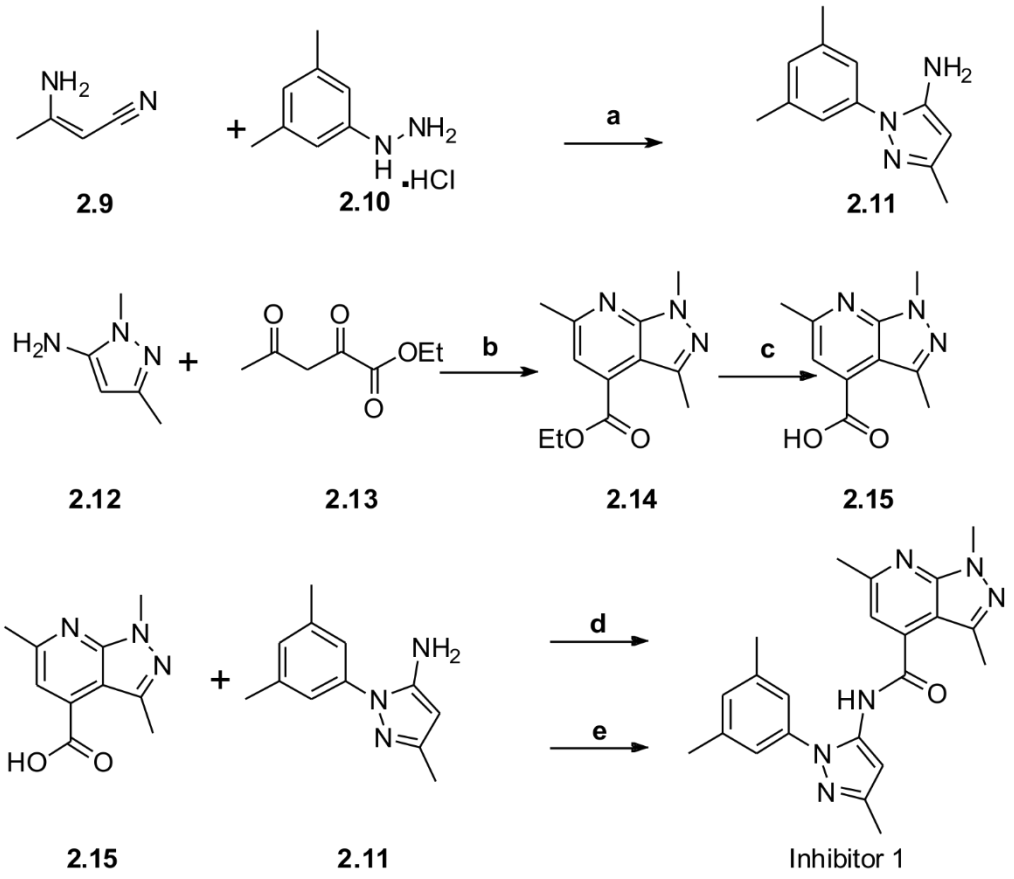
The results from enquiry six led to the selection of three target compounds for further computational analysis with the NCS1 D2R peptide system, synthesis (Section 2.2) and biophysical analysis (Section 2.3).

2.2 Synthetic Approach

From the computational design processes, three compounds were selected for synthesis. The synthetic routes to inhibitors 1, 2, and 3 were selected after a rigorous retrosynthetic investigation into possible routes in order to efficiently produce the desired compounds.

2.2.1 Synthesis of Inhibitor 1

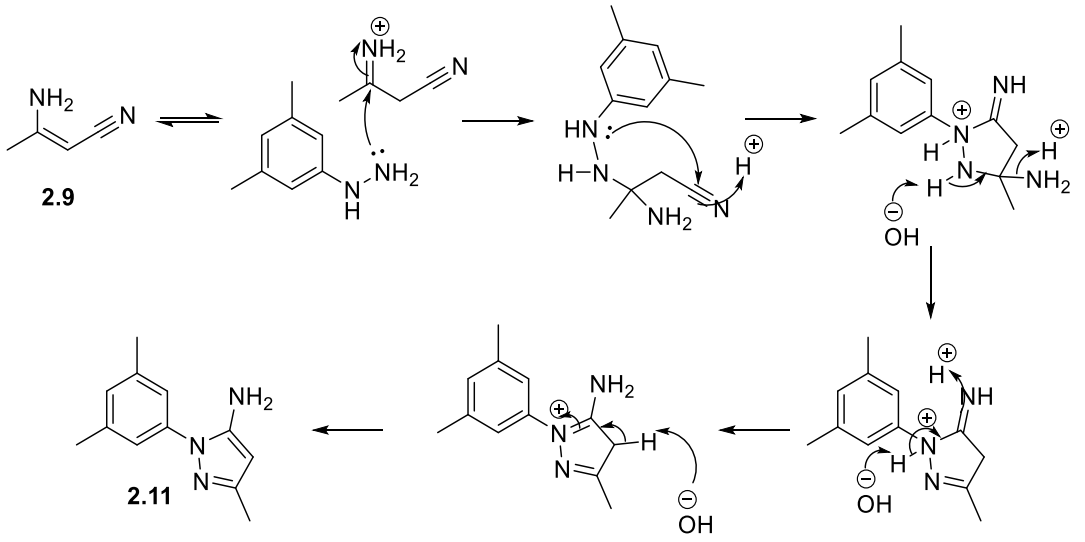
The first compound selected for synthetic investigation, Inhibitor **1** was chosen due to its predicted ability to mimic four of the six desired pharmacophoric points (Compound **2.7** Table 2.9). The synthesis of Inhibitor **1** can be seen in Scheme 2.2 and involved four steps to achieve the desired compound.



Scheme 2.2- **Synthetic route to Inhibitor 1**. **(a)** HCl, 115°C 4 hrs, 25 °C, NaOH, 80%. **(b)** AcOH, 120 °C 4 hrs, 83%. **(c)** *i*-PrOH, KOH, 83 °C 5 hrs, 25 °C, 2 M HCl, 72%. **(d)** *N-N* diisopropylethylamine, 1-methyl-2-chloropyridinium iodide, anhydrous MeCN, 83 °C, 4 hrs 44%. **(e)** DCC, HOBt, anhydrous THF 25 °C, 30 min, 25 °C (24hr), 50 °C (24hr), 28%.

The first step in the synthesis involves a condensation reaction between 3-aminocrotonitrile **2.9** and 3, 5-dimethylphenyl hydrazine hydrochloride **2.10**. The reaction was initially heated to 115 °C in a solution of 1 N HCl for four hours; basification of the reaction with the drop wise addition of 6 N NaOH at room temperature initiated cyclisation to form the amino pyrazole **2.11** (Scheme 2.3). The

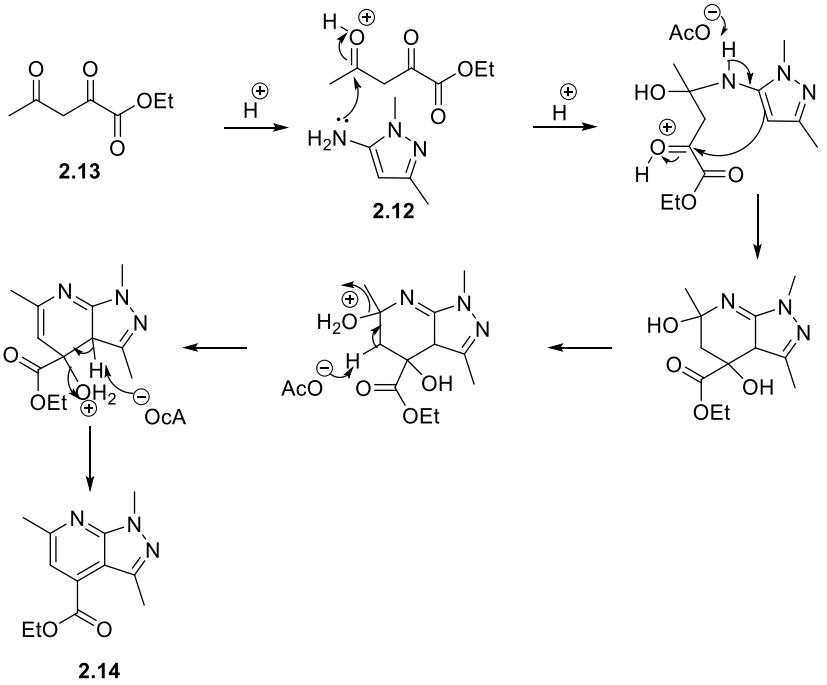
reaction afforded the desired product **2.11** with a good yield of 80% following modification of the published work up and isolation method.[245]



Scheme 2.3– The proposed mechanism for the condensation and cyclisation reaction between 3-aminocrotonitrile and 3, 5-dimethylphenyl hydrazine hydrochloride.

The consecutive Combes-type condensation reaction (Scheme 2.4), between ethyl-acetopyruvate **2.13** and 1,3 -dimethyl-1H-pyrazole-5-amine **2.12** was achieved following a procedure adapted from a reaction developed by Volochnyuk.[246] Combes reactions are used extensively in the synthesis of quinolines and typically involve a condensation reaction between an un-substituted aniline and a β-diketone, followed by an acid catalysed ring closure of the intermediate imine or enamine.[247] The procedure itself was straightforward and involved heating the reagents in acetic acid to reflux at 118 ˚C for four hours; however the initial attempt at this reaction gave the product in low 19% yield. This was believed to be in part due to the workup procedure, in which the solvent was removed *in vacuo*, the resultant residue triturated with water and recrystallized from *iso*propanol (*i*-PrOH).[246] Therefore repeating the procedure altering to the work up procedure, removing the trituration and recrystallization steps, replacing them with an organic extraction into dichloromethane (DCM) and a water wash. This simple change substantially increased the yield achieving 83% of the desired pyrazolepyridine ester (**2.14** Scheme 2.2). Subsequent saponification of the ester **2.14** to the carboxylic acid derivative **2.15** was initially low with a recovery of around 33% of the desired product. Once again this was attributed to the workup of the reaction, which was modified to organic extraction and led to an improved yield of 72%. It should be noted that despite the improvement in the yield this method was not completely effective; the reaction was partially insoluble in organic solvent and so formed an emulsion at the organic/ aqueous interface which

was believed to lead to a slight loss of product. However, the high solubility of the inorganic side products in the aqueous phase allowed for the isolation of the acid without need for further purification using flash column chromatography.



Scheme 2.4- Proposed mechanism for the formation of ester **2.14.**

The final step in the synthesis of Inhibitor **1** was the amide coupling reaction between the amine **2.11** and the carboxylic acid **2.15**. The first reaction followed a procedure developed by Volochnyuk, using *N,N'*- diisopropylethylamine (DIPEA) as the organic base and 1-methyl-2-chloropyridinium iodide also known as "Mukaiyama Reagent" as the carboxylic acid activating agent.[246,248] The reaction was successful in yielding the desired amide Inhibitor **1**, with a low yield of 23%. Proton NMR experiments indicated the possible presence of two rotameric forms of the product (Figure 2.14), this was observed through the presence of second peak for each hydrogen environment, with the integration of one eighth that of the parent peak. This particular reaction was not investigated further, however temperature controlled NMR experiments could be used to confirm this.

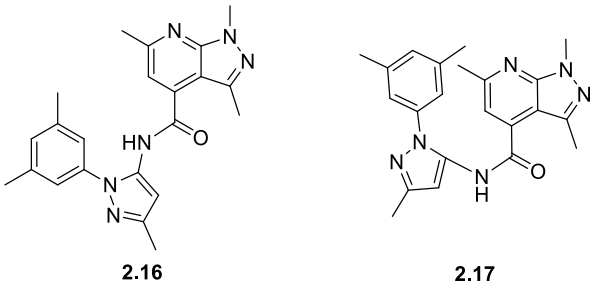
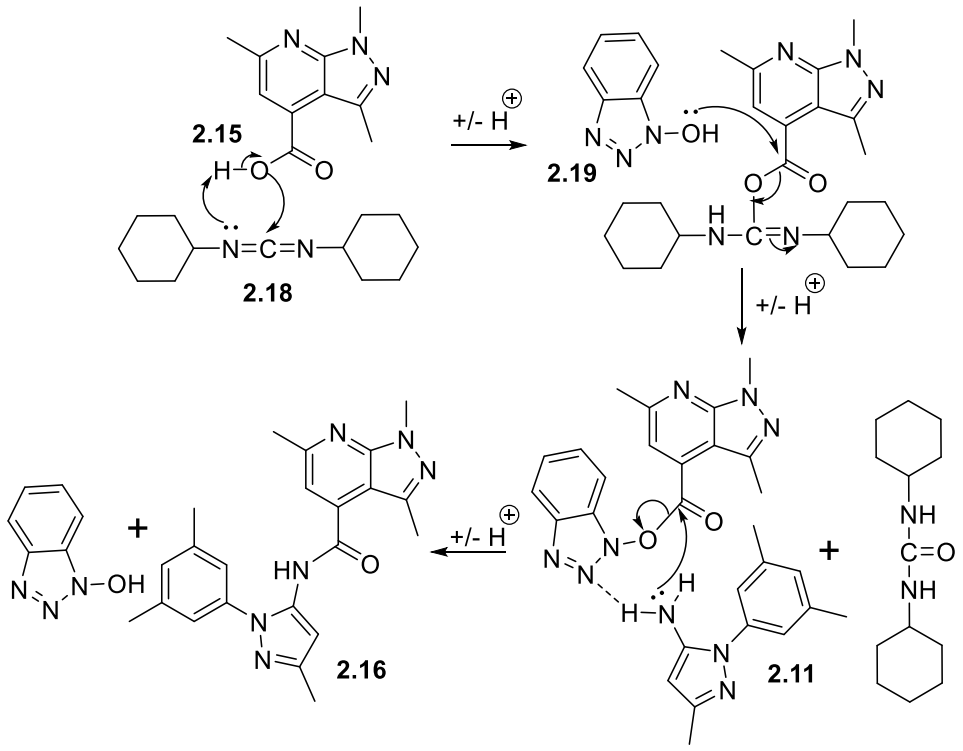


Figure 2.14- Two proposed rotameric forms of Inhibitor1 **2.16** is the most favoured over **2.17.**

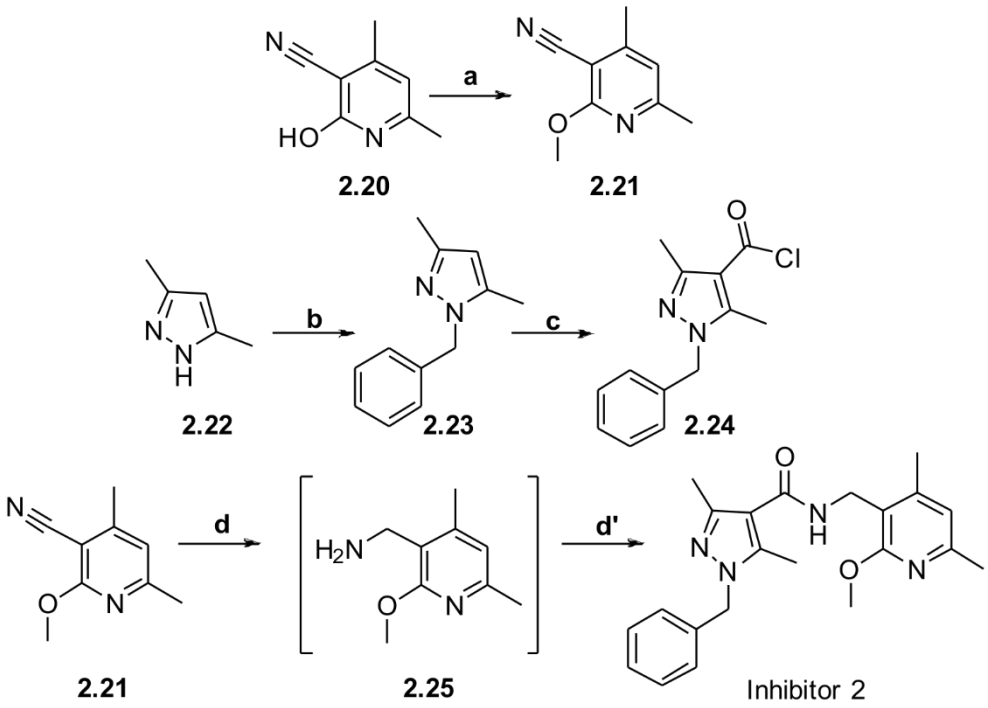
The second methodology investigated, involved using *N, N'*-dicyclohexylcarbodiimide **2.18** (DCC), a coupling reagent that belongs to a wider family of coupling agents known as the carbodiimides and the additive hydroxybenzotriazole **2.19** (HOBt). The general mechanism involves reaction of the carboxylic acid with the carbodiimide, forming the *O*-acylurea intermediate which can then react with a number of nucleophiles such as the amine to form the desired product.[249] However, problems with this mechanism include the formation of unwanted *N*-acyl urea side products, which can significantly reduce the reactivity of these reagents. Therefore, additives such as HOBt or 4-dimethylaminopyridine (DMAP) are used to prevent such side reactions from occurring.[249,250] They work by reacting it with the *O*-acylurea to give, in the case of HOBt, the hydroxybenzotriazole activated ester, which is thought to be more reactive towards amines, as it stabilises the approach of the amine through a series of hydrogen bonds. The proposed mechanism of the reaction can be seen in Scheme 2.5.[250] Inhibitor 1 was isolated with a yield of 28% and proton NMR indicated the presence of a single isomer assumed to be the more favoured product (**2.16** Figure 2.14) and this material was used *in vitro* binding studies (Section 2.3).



Scheme 2.5- The proposed mechanism of a DCC **2.18** and HOBt **2.19** peptide coupling reaction, to form **2.16**.

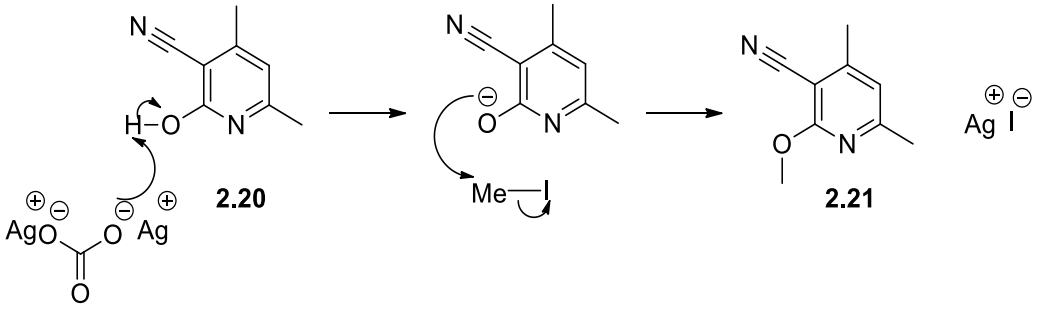
2.2.2 Synthesis of Inhibitor 2

The second compound selected for synthesis was Inhibitor 2 (Scheme 2.6), the original scaffold selected from the screening procedures contained a chlorine group on the pyrazole ring (Compound **2.8** Figure 2.9). However, for the purposes of the synthesis it was modified, such that the chlorine group on the pyrazole ring was replaced with a second methyl group. This was selected due to the commercial availability of the pyrazole compound.



Scheme 2.6- Synthetic route to Inhibitor 2. **(a)** DCM, Ag_2CO_3, MeI 71%. **(b)** DMSO, KOH 80 °C 1 hr, $PhCH_2Br$ 2 hrs 25°C 70%. **(c)** Oxalyl chloride 63°C reflux 3 hrs. **(d)** $LiAlH_4$, anhydrous Et_2O, 2 hrs, 25 °C, 2 hrs reflux, d' NaOH 20%, **2.24**, pyridine, 24 hrs 25°C 32%.

In the synthetic route for Inhibitor **2** (Scheme 2.6), the first step of the synthesis was methylation of 2-hydroxy-4,6-dimethylnicotinonitrile **2.20**. The procedure was originally carried out by Wijtmans *et al.* on a similar starting material 3-bromo-4,6-dimethylpyridone as a part of a larger synthesis of a final compound a substituted pyrindol.[251] The reaction involved stirring the reagents in DCM at 25 °C for twenty four hours with protection from light sources to prevent the degradation of the sensitive silver carbonate Ag_2CO_3. After several attempts use of fresh reagents gave a reasonable 71% yield of the methoxypyridone **2.21**.



Scheme 2.7- Proposed methylation reaction of **2.20** to form **2.21.**

Benzylation of the pyrazole **2.22** was carried out using benzyl chloride in a solution of potassium hydroxide KOH and DMSO. Benzylation reactions of unsymmetrical pyrazoles tend to be unselective for the nitrogen which reacts if an inorganic base is used, generating a mixture of isomers.[252] This was not an issue with the symmetry involved and pyrazole **2.22** was *N*-benzylated with a pleasing yield of 70%. The subsequent formation of the acid chloride **2.24**, was performed through the reaction of compound **2.23** with refluxing oxalyl chloride **2.26**. Chiriac *et al.* indicated that the mechanism of the reaction occurred *via* the oxalyl intermediate, which could then eliminate a CO to form a ketene intermediate. Finally this is then attacked by chloride giving the acid chloride product (Scheme 2.8).[253] The product formation was followed by thin layer chromatography (TLC) and upon completion was not purified as it was believed that any remaining intermediate could be removed after the final coupling reaction. Despite this, the crude product was analysed using infrared spectroscopy which indicated the presence of the carbonyl stretch at 1735 cm^{-1} confirming the acid chloride **2.24** had successfully formed.

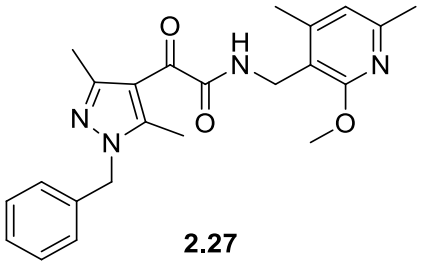
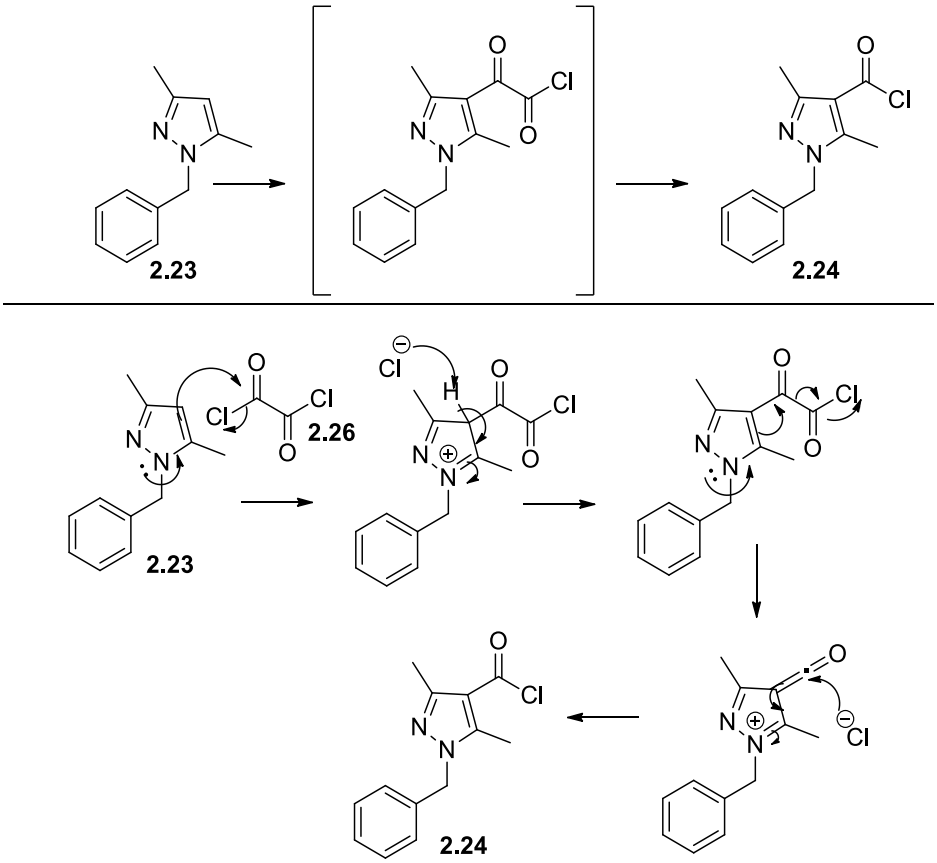


Figure 2.16 - Proposed major undesired oxalyl side product **2.27.**

The final step in the synthesis combined two separate components in a one pot reaction (Scheme 2.6). Initial reduction of the cyano group contained within **2.21** to the corresponding amine intermediate **2.25** was conducted using a $LiAlH_4$ slurry in dry ether, the reaction was stirred for two hours at 25 °C and then heated to reflux for a further two hours. Decomposition of the hydride was then carried out using the minimum volume of 20% aqueous NaOH. Upon completion of the formation of the

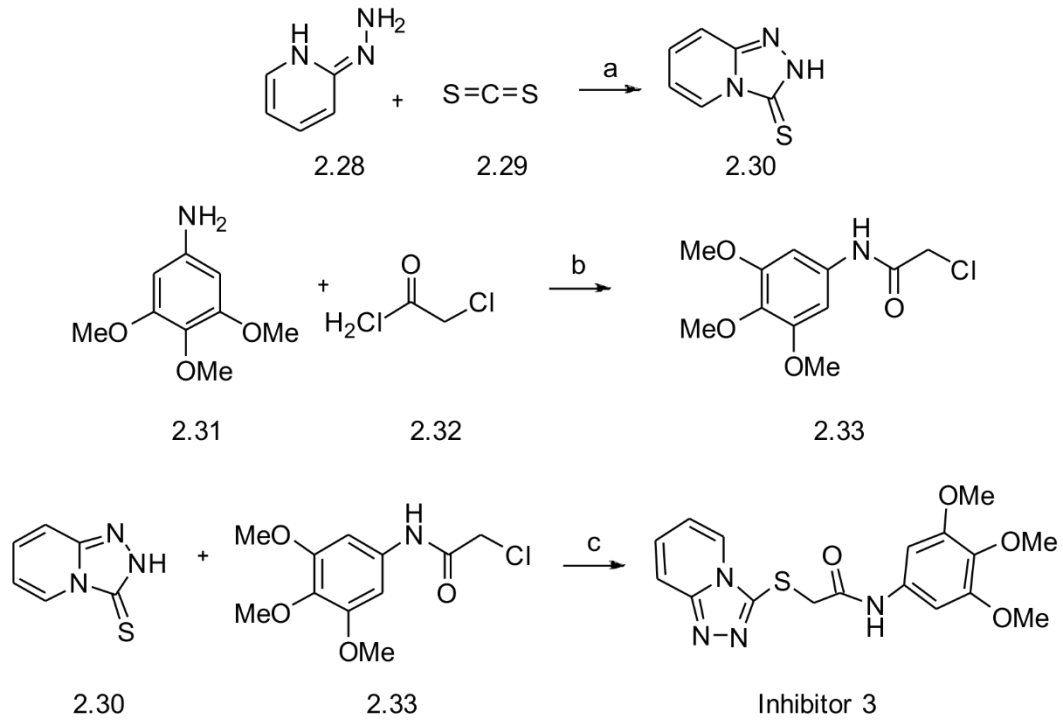
amine intermediate, the crude acid chloride **2.24** was added dissolved in the minimum volume of dry ether, along with pyridine. The reaction was then left to stir overnight at room temperature and after purification gave Inhibitor **2** in a slightly lower than expected yield of 30%. This can be attributed to the formation of a major undesired side product (**2.27** Figure 2.16) believed to form through reaction with the oxalyl intermediate of the acid chloride reaction (Scheme 2.8) and the amine **2.25**. This was confirmed by accurate mass spectrometry which indicated the presence of a second carbonyl group within the molecule. The proton NMR did not indicate any change in the hydrogen environments; however an additional keto carbonyl peak in the carbon NMR spectrum (around 186 ppm) further confirmed the major product of the reaction was in fact the oxalyl amide (**2.27** Figure 2.16). This implies that the acid chloride should in future be purified to prevent a side reaction occurring between the intermediate oxalyl chloride and the amine **2.25**. Inhibitor **2** was tested *in vitro,* results of which can be found in Section 2.3, compound **2.27** however, was not tested using biophysical techniques.



Scheme 2.8- Proposed mechanism of acid chloride formation *via* an oxalyl derivative.

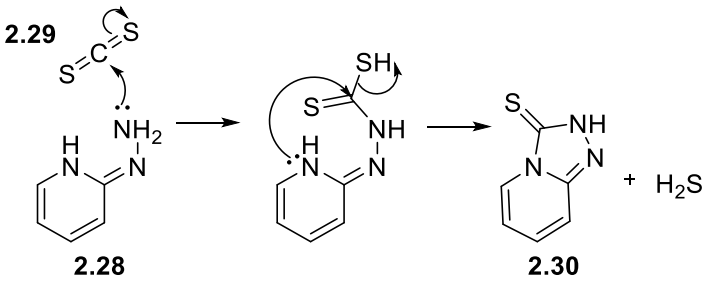
2.2.3 Synthesis of Inhibitor 3

The final compound selected from the computational design process was Inhibitor 3, which was predicted as being able to mimic three of the desired six pharmacophoric points (Compound **2.5,** Table 2.9). As with Inhibitor **2**, due to the commercial availability of the precursor compound **2.28**, the final target compound was modified such that the two *meta*-chloro groups on the pyridine fragment (Compound **2.5**, Figure 2.9) were removed.



Scheme 2.9- Synthetic route to Inhibitor 3. (a) CS_2**,** $CHCl_3$, 2-hydrozinopyridine, reflux 20 hrs, 76% **(b)** 3,4,5-trimethoxyaneline, chloroacetyl chloride, triethylamine, reflux 15 hrs, 84%. **(c)** *N*-*N*-diisopropylethylamine, anhydrous DMF, 120 ˚C, 1 hr, 65%.

Inhibitor **3** was synthesised using a succinct route that involved only three steps (Scheme 2.9). Firstly, 2-hydrazinopyridine **2.28** in chloroform was heated to reflux at 62 ˚C along with carbon disulfide **2.29** for 20 hours, forming the known triazolopyridinethione **2.30**. This compound had been used previously as a fragment in an SAR study for hit to lead development of anti-malarial agents.[254] The reaction mechanism is a cyclocondensation (Scheme 2.10) and afforded the product with a yield of 76%. This was followed by *N*-acylation of 3,4,5-trimethoxy aniline **2.31** with chloroacetyl chloride **2.32** to form the corresponding amide **2.33** in 84% yield.



Scheme 2.10- Proposed cyclocondensation reaction mechanism between 2-hydrazinopyridine and carbon disulphide.

The final step in the formation of Inhibitor **3** was the coupling reaction between the thione **2.30** and the acetamide **2.33**, achieved by heating both precursors in anhydrous DMF for 1 hour. The resultant crude product was purified using flash column chromatography to afford the final compound in 65% yield. The Inhibitor was then tested *in-vitro* (Section 2.3).

2.2.4 Summary of Synthesis

The synthesis of the first generation inhibitors was successful in efficiently producing three diverse unrelated ligands. These small molecules are designed to mimic the 3-D pharmacophore and to disrupt the protein-protein interaction between NCS1 and D2R peptide. Inhibitors **1** and **2** were synthesised in a reasonable yield, 28% & 32% respectively, in four or five synthetic steps respectively. Inhibitor **3** was produced in an overall 65% yield in a three step procedure. The ease of the synthesis would facilitate synthesis of analogues following *in vitro* binding. Inhibitors **1**, **2** and **3** were taken forward for biophysical analysis to determine their binding interactions with NCS1.

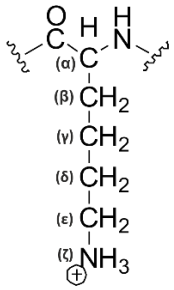
2.3 Binding assessment using Biophysical Methods

In the design process for small molecule modulators of PPIs it is important to understand how the designed ligand interacts with the protein of interest. To do this there are a number of biophysical techniques which can be employed such as; X-ray crystallography; fluorescence polarisation (FP); nuclear magnetic resonance (NMR) spectroscopy; isothermal titration calorimetry (ITC) and intrinsic fluorescence spectrophotometry. It is the final three that were used in this research. Each of these techniques were mentioned Section 1.6 of the Chapter 1 and will be described in greater detail in this chapter, with specific reference to the experiments used. A detailed discussion of the results from the biophysical characterisation of the first generation of compounds will also be given 2.3.2.

2.3.1 Introduction to protein assignment

From basic biochemistry we know that proteins are composed of 20 different amino acids, although the number and composition of amino acids in each protein differs substantially. The first step in the NMR study of protein structures and interactions is the assignment of NMR resonances to specific atoms of individual amino acids within the polypeptide chain of a protein. To do this we employ a number of 2-D HSQC and 3-D experiments that are able to determine the backbone and sidechains resonance assignments for different atoms in each amino acid residue within that protein. As introduced in Chapter 1.6 the chemical shift (δ) for each of these atoms and therefore residues will be slightly different to the other as each shift is dependent on the chemical environment that surrounds it.

In essence by employing NMR experiments such as the 2-D ^{1}H ^{15}N HSQC spectra of the highly sensitive NH atoms with other 3D experiments, we can obtain a fingerprint for our proteins that informs us how the residues are arranged. The assignment can subsequently be used to inform us of the amino acid residues effected by the interactions of small molecules with the protein, this is conducted using chemical shift mapping (discussed in section 2.3.3 Introduction to ligand binding studies). To understand backbone and sidechain assignment we must be aware of the basic nomenclature of the atoms that comprise each amino acid (Figure 2.17 and 2.18).



Lysine (k)

Figure 2.17- Amino acid nomenclature with respect to the sidechain of the amino acid lysine, the main chain is attached to the α CH.

The backbone of a protein is composed of repeating units of amino acids which are made up of an NH group followed by a carbon atom C_α, which is attached to the carbonyl CO, in the amino acid these form a chain *via* the peptide bond joining the CO of one amino acid to the NH of the adjacent amino acid this continues in this way along the protein sequence (Figure 2.18). The differences between amino acids occur at the sidechain which is connected to the backbone through the C_α, an example of a sidechain and its nomenclature for the amino acid lysine can be seen (Figure 2.17). For the purposes of assignment the amino acid sequence is described using the terms i residue and i-1, meaning the current residue and the one which precedes it. The backbone assignment utilise pairs of 3-D experiments whose magnetisation pathways are described in reviews ([255–261]).

The methodology for backbone resonance assignment is sequential and occurs *via* the connections between HSQC-based experiments to link NH atoms of adjacent amino acids together through C_α correlations (in the case of CbCa(CO)NH and HNCaCb experiments Figure 2.18). This can also be conducted through the CO correlations (HNCO and the HN(Ca)CO experiments Figure 2.19) finally, the CO of each amino acid can be correlated with the H_α and hence H_β (HNCa and HN(CO)Ca experiments, Figure 2.20).[259,261]

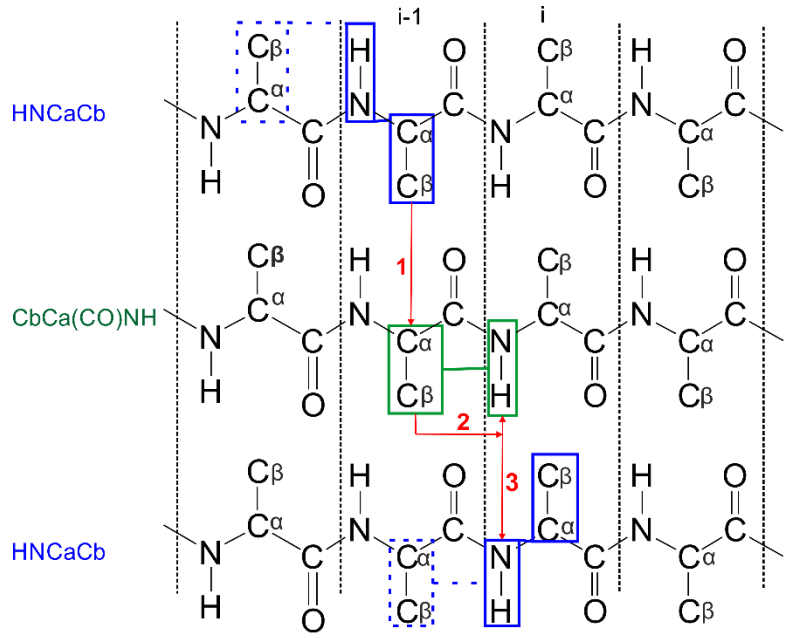


Figure 2.18- The sequential backbone assignment method used for the assignment of NCS1 using two 3-D experiments, the CbCa(CO)NH and HNCaCb. The HNCaCb experiment is used to 'connect' the NH resonance with its own $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ (seen in the solid blue lines) and to the $^{13}C_{a}$ and $^{13}C_{b}$ of the preceding residue (dotted blue lines). The HNCaCb can then be correlated to the respective peak found in the CbCa(CO)NH (red pathway 1) which links a NH group of a residue to the $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ of the residue that precedes it (both seen within the green boxes). This linking process allows for the identification of the (i -1) residue in front of the (i) residue in question. In essence walking us along the protein backbone and linking each residue in the repeated consecutive pathway (seen in red steps 2 and 3 etc.).

In each case, the NH correlates to the preceding residues carbon and only in the HNCaCb, HN(Ca)CO, HNCa and HN(CO)Ca does the NH also correlate to the intra-residual carbon, thus allowing for differentiation of intra and inter-residual correlation in order to assign each amino acid in the sequence.

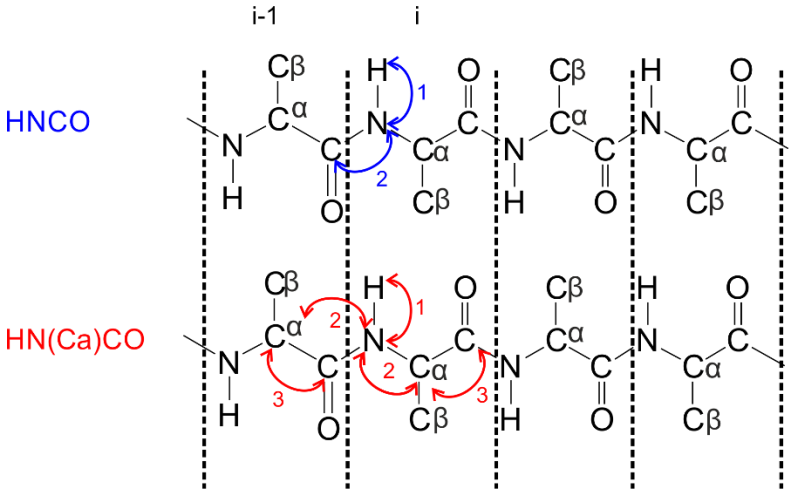


Figure 2.19- The transfer of magnetisation for the experiment HNCO (blue), the magnetisation originates on the ^{1}H and is passed to the attached ^{15}N (pathway 1), the ^{15}NH ^{13}CO selectively transfers the magnetisation to the CO (pathway 2) of the residue and is subsequently transferred back for detection. The HN(Ca)CO (red) where the magnetisation is transferred from ^{1}H to the ^{15}N (pathway 1) and then $^{13}C_{\alpha}$ (pathway 2) finally it is transferred to the ^{13}CO (pathway 3). The process is then reversed as the magnetisation is transferred back to the ^{1}H for detection.

Using these three different pairs of experiments allow us to follow the chain of amino acids along the backbone of the protein and provide us with full NMR backbone assignment. It should be noted that prolines are unobservable, hence they often appear as a natural break along the chain.

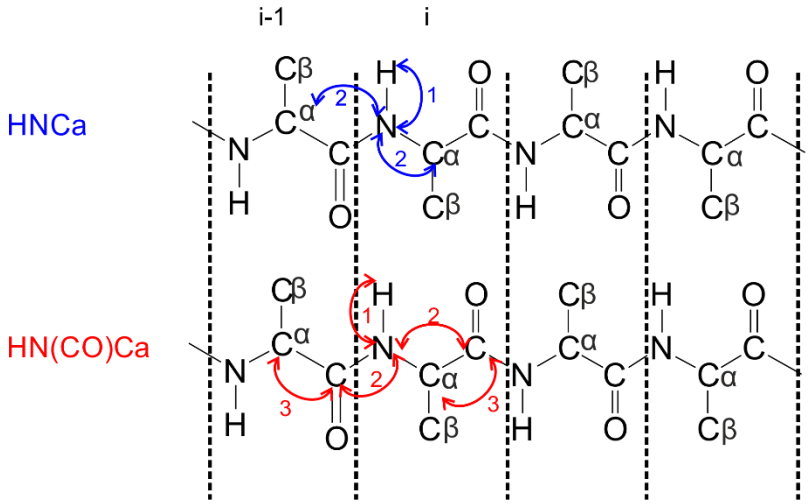


Figure 2.20- The transfer of magnetisation that occurs in the HNCa the magnetisation is transferred from ^{1}H to the ^{15}N (pathway 1) and then $^{13}C_{\alpha}$ (pathway 2) and subsequently transferred back for detection. As with the coupling of ^{15}NH group in the NH(Ca)CO experiment, the ^{15}NH group is also coupled to the $^{13}C_{\alpha}$ of its own residue i and that of the preceding i-1. In the HN(CO)Ca experiment the magnetisation is transferred from the ^{1}H to the ^{15}N (pathway 1) to the ^{13}CO (pathway 2) and then onto the $^{13}C_{\alpha}$ (pathway 3), where it is then transferred back along the same path with no evolution of signal relating to the ^{13}CO.

To use the NMR data for investigative studies of ligand interactions it is important to note that complete backbone assignment should be accompanied by assignment of the side-chain residues. To assign the side-chain of each residue, 3-D HC(C)H- total correlation spectroscopy (TOCSY) NMR experiments are used in conjunction with the previously collected backbone data.

In a 3-D HC(C)H - TOCSY experiment, magnetisation is transferred from the sidechain hydrogen nuclei onto the attached ^{13}C nuclei, where there is an isotropic ^{13}C mixing stage and the magnetisation is then transferred back to the sidechain hydrogens, hence the HC(C)H. This experiment provides the carbon and hydrogen chemical shifts of the side-chains for most of the residues including Hγ, δ and Cγ, δ, ε (when the atoms are connected through the C_{β} of a straight chain and excludes quaternary carbon connections). Using the previously acquired Hα, β chemical shifts of each residue, it is possible to link the corresponding Hγ, δ, ε for that particular residue through TOCSY and hence assign the corresponding Cγ, δ, ε chemical shifts.

2.3.2 Backbone and sidechain assignment of NCS1

General experimental procedures

For NMR investigation all NMR spectra were acquired at 298K unless otherwise stated, on either a Bruker AVANCE II+ 600 MHz or 800 MHz spectrometers, each equipped with a 5 mm TCI triple-resonance cryoprobe. The NMR samples were prepared to a volume of 550 µL including 10% (v/v) [$^2H_2$]O, the NMR buffer used was composed of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$, pH 6.4.

^{15}N labelled and ^{13}C ^{15}N labelled NCS1 were prepared as will be described in detail in Chapter 6. Briefly recombinant NCS1 was obtained from an expression construct containing the *Rattus Norvegicus* NCS1 gene, [262] inserted into a pETM-11 vector and subsequently transformed into BL21 DE3 E-coli competent cells (Novagen).[263] The *N*-terminally His_6-tagged protein was purified using nickel affinity chromatography and size exclusion chromatography and the purified protein solution was buffer exchanged into NMR buffer using a Sephadex G-25 Medium Gravity-Flow PD-10 (GE Healthcare) column that had been pre equilibrated with the NMR buffer.

Data collected was processed using Topspin 3.1 (Bruker) and the Azara processing package which is contained within the CCPN analysis software. The latter software was also used for resonance assignments. Further experimental details can be found in Chapter 6 Section 6.3.2.

2.3.2.1 Results and Discussion

For the purposes of assigning the protein backbone and sidechain residues, triple resonance assignment was conducted using 2-D ^1H ^{13}C and ^1H ^{15}N HSQC spectra, which were used alongside specific pairs of experiments designed for backbone and sidechain assignment.[264] These experiments included CbCa(CO)NH and HNCaCb, HNCa and HN(CO)Ca, HNCO and HN(Ca)CO (see 2.3.1 for a full explanation of the experiments). Assignment of the sidechain residues was conducted using 2-D ^1H ^{13}C HSQC and HC(C)H TOCSY experiments (Figures 2.24, 2.25, 2.26 and Appendix Tables A.2.1 and Table A.2.2) in the desired NMR buffer.

A total of 91% of backbone atoms and 42.7% of the sidechain resonances were assigned. The final assigned 2-D ^1H ^{15}N HSQC backbone spectra can be seen (Figure 2.22 and 2.23), alongside the assigned 2-D ^1H ^{13}C aliphatic and aromatic methyl sidechain spectra (Figure 2.24 and 2.26).

The secondary structure of NCS1 predicted from the backbone assignment and chemical shift index (CSI) [265] was then compared to that of the known structure PDB (protein data bank) code 5AER and can be seen (Figure 2.21), indicating the correct folded form of unbound NCS1.[266] It can be seen that there are 10 α helices which form the four EF hands predicted using CCPN analysis and subsequently determined through crystallographic data, each pair of EF hands are connected by a hinge loop.[13,267]

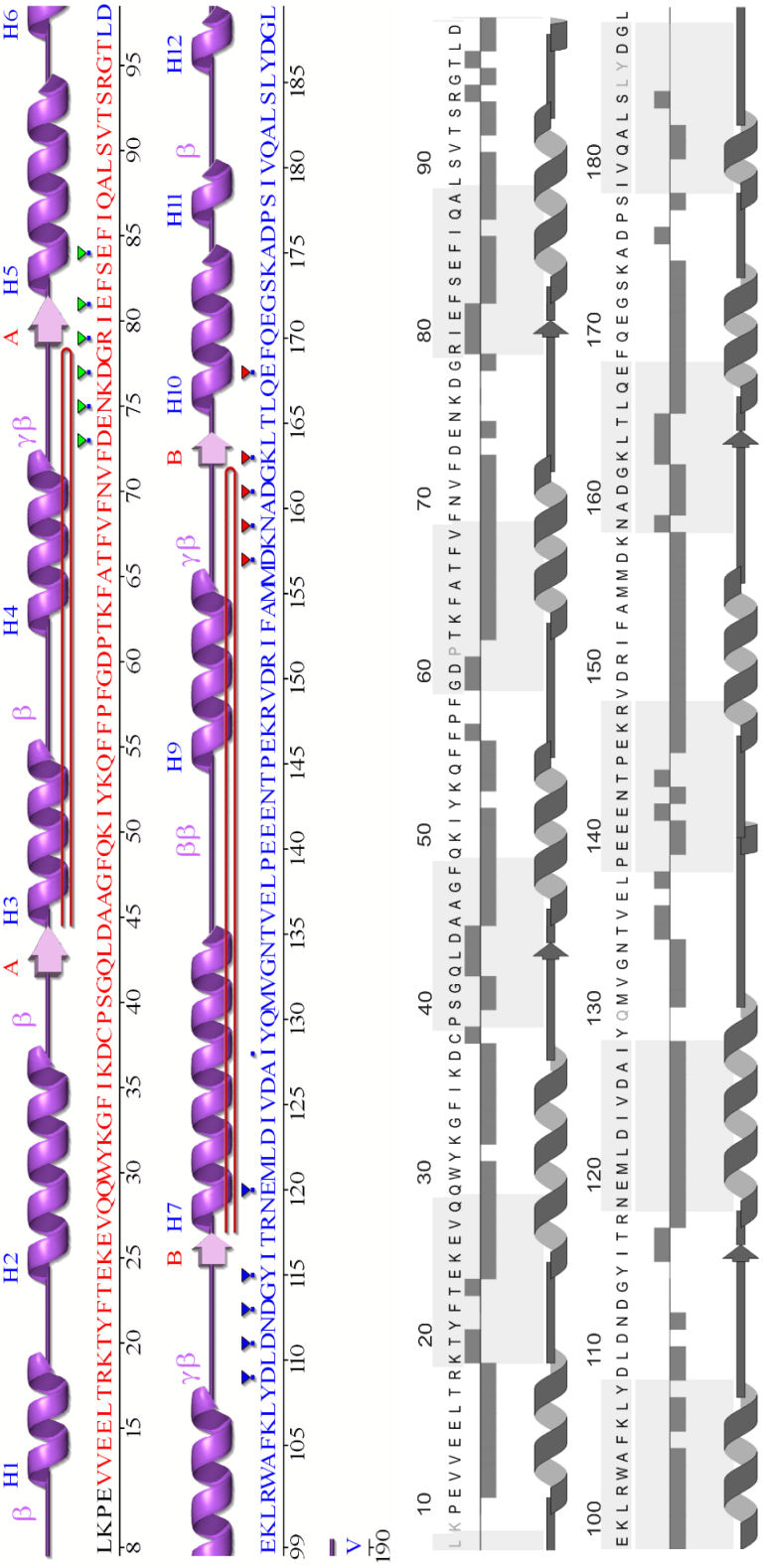


Figure 2.21- Calculations of the secondary structure of NCS1. **(a)** The known secondary structure of neuronal calcium sensor 1 (NCS1) from *rattus norvegicus* (PDB 5AER) through crystolographic techniques with a 2.19 Å resolution[13,267]. **(b)** The calculated secondary structure of NCS1 determined using the NMR backbone assignment and CSI prediction.

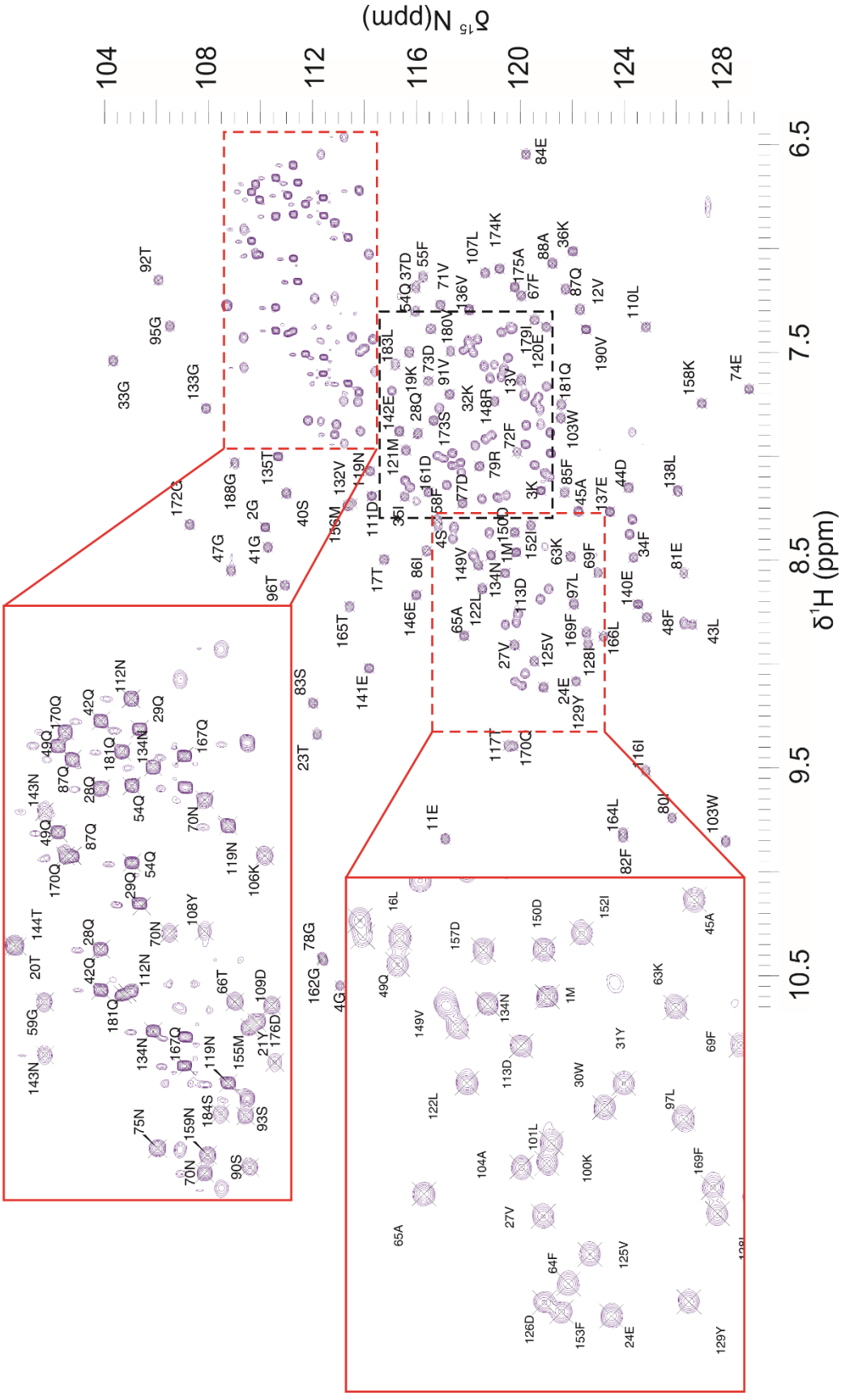


Figure 2.22- Assigned backbone and sidechain spectra of NCS1 (Part i). A 2-D ^{1}H ^{15}N HSQC spectra of ^{15}N ^{13}C NCS1 0.9 mM acquired at 300 K in NMR buffer pH 6.4 on Bruker AVANCE III 800MHz spectrometer. All known main chain sidechain resonances are labeled in black, those peaks within the black dashed box are highlighted in the following figure.

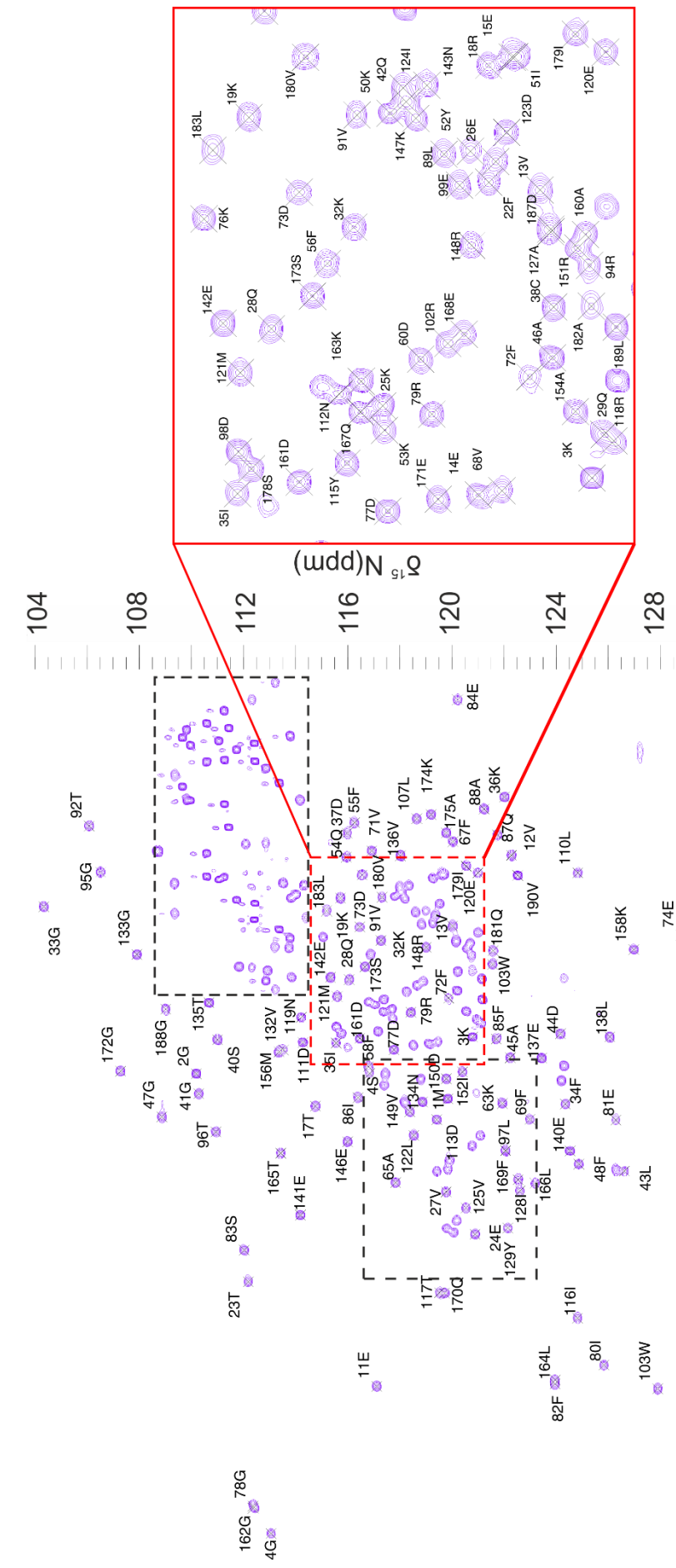


Figure 2.23- Assigned backbone and sidechain spectra of NCS1 (Part ii) A 2-D ^{1}H ^{15}N HSQC spectra of ^{15}N ^{13}C NCS1 0.9 mM acquired at 300 K in NMR buffer pH 6.4 on Bruker AVANCE II 800MHz spectrometer. Continuation of the assigned backbone and sidechain spectra of NCS1 Figure 2.22 Part i, with those residues within the black dashed box highlighted in the previous figure.

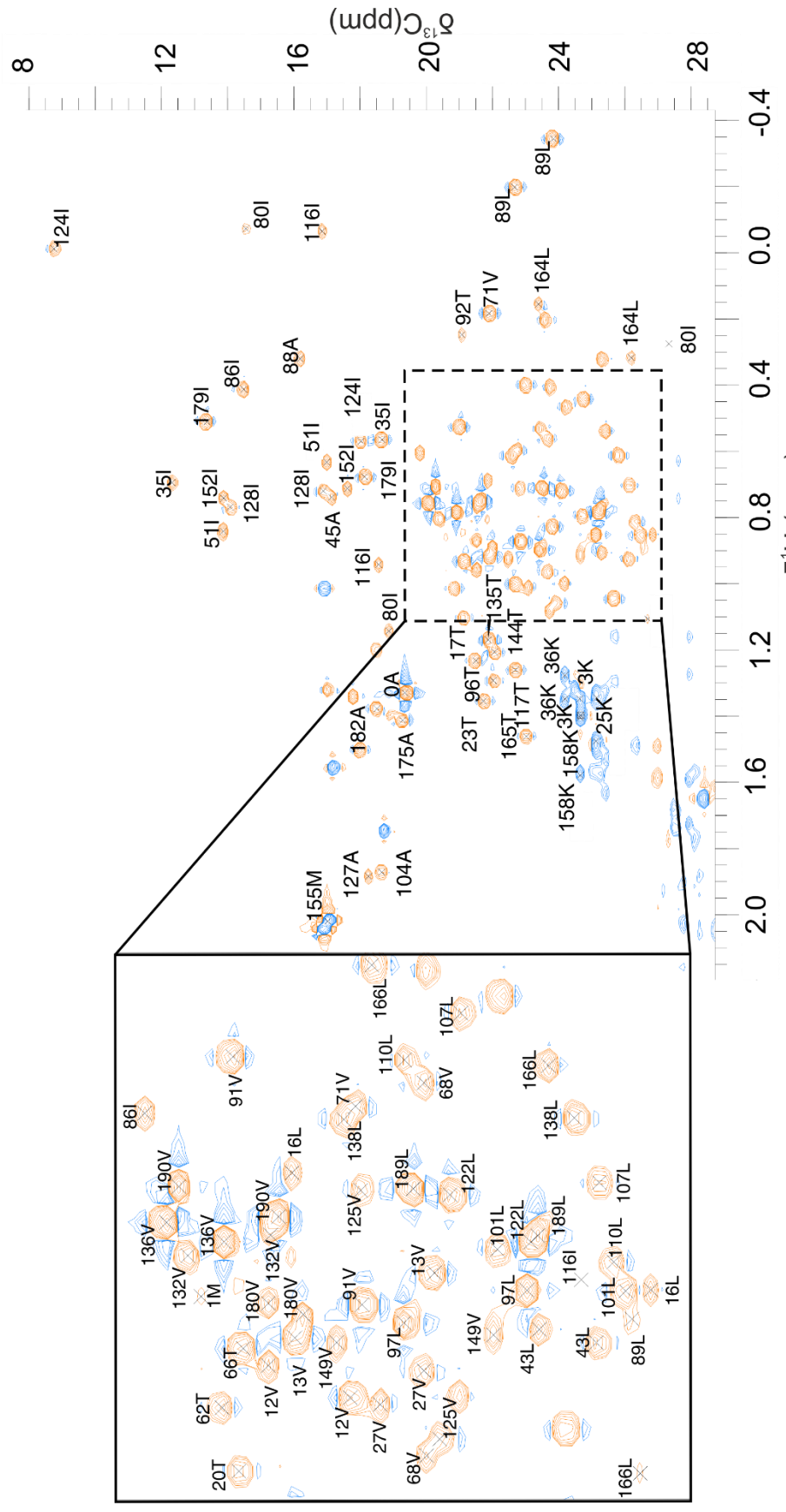


Figure 2.24- Assigned methyl sidechain spectra. The enlarged methyl sidechain region of ^{15}N ^{13}C NCS1 50 µM taken from ^{1}H ^{13}C HSQC acquired at 298K in NMR buffer pH 6.4 (see Section 2.3.2 for further details) on Bruker AVANCE II 600MHz spectrometer. Those residues for which assignment was possible are labelled to the related resonance in the spectra.

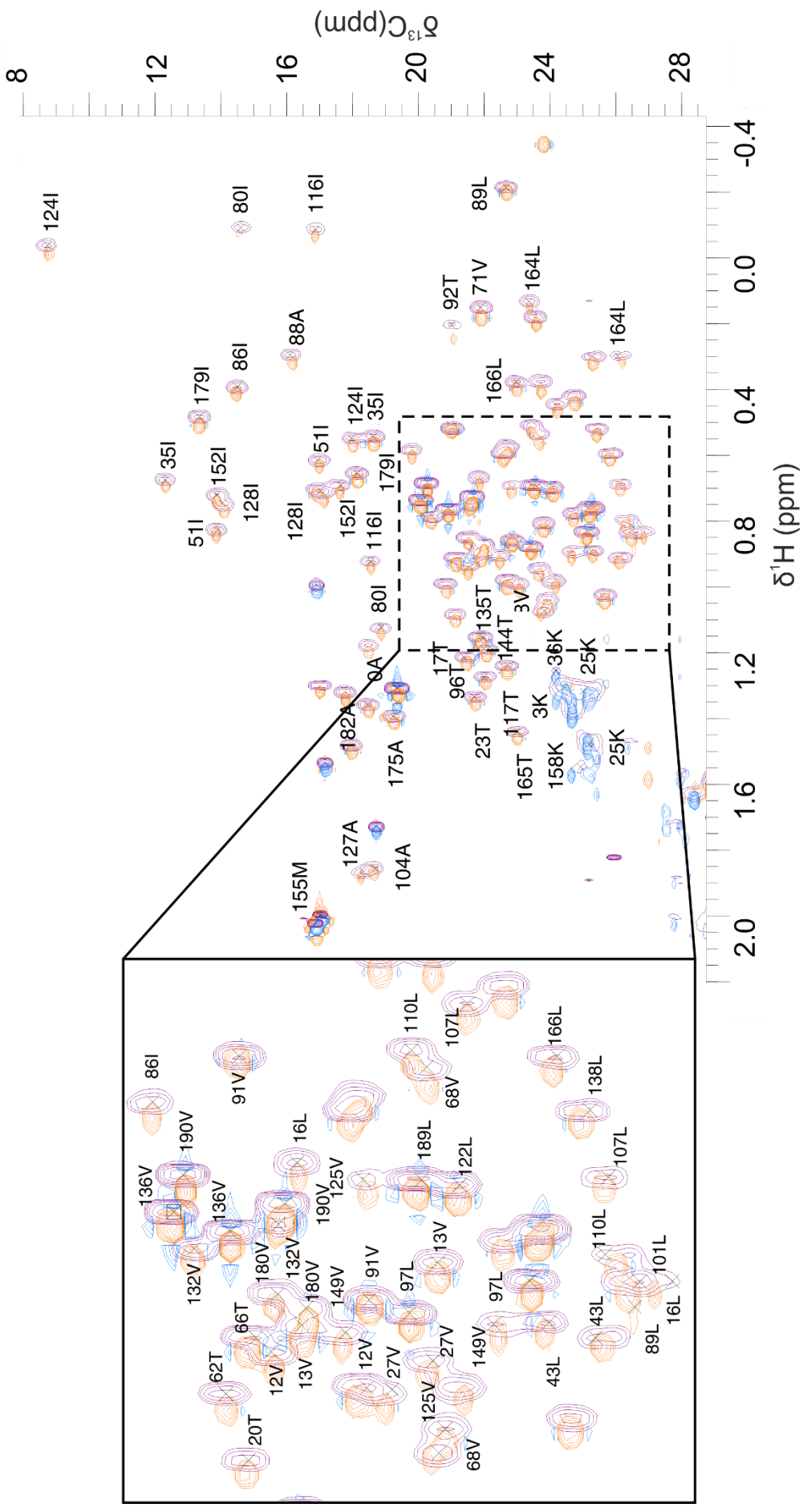


Figure 2.25- An overlay of two methyl sidechain spectra of NCS1 with and without DMSO. The ^{1}H ^{13}C HSQC spectra of ^{15}N ^{13}C NCS1 50 µM orange and blue overlaid with the spectra of ^{15}N ^{13}C NCS1 50µM and 0.5% DMSO purple. Those residues for which the transfer of assignment was possible can be seen. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

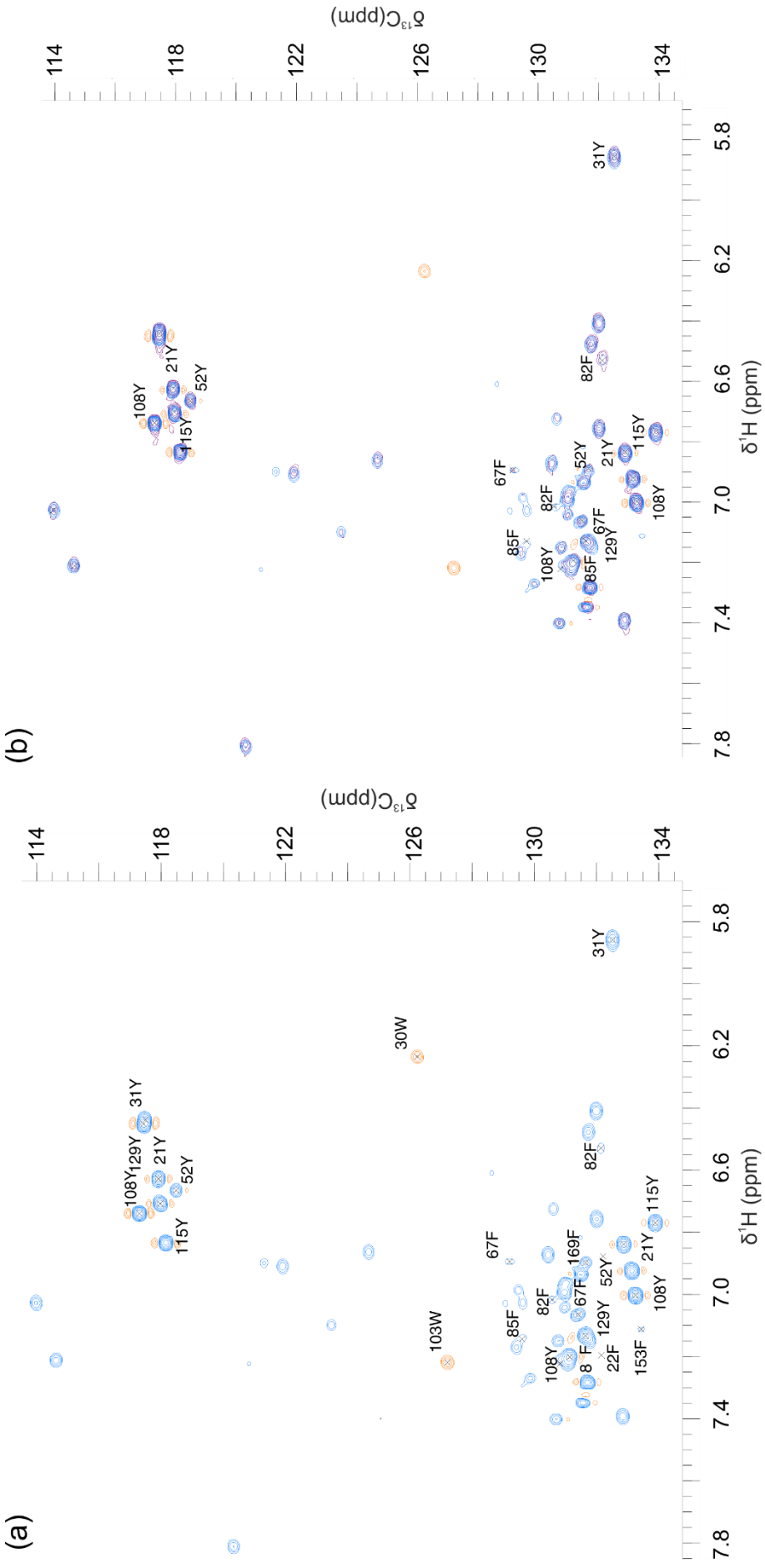


Figure 2.26- Two aromatic sidechain spectra of NCS1 with and without DMSO. **(a)** ^{15}N ^{13}C NCS1 50 µM with the assigned aromatic sidechains of the amino acid residues labelled. **(b)** ^{15}N ^{13}C NCS1 50 µM Blue and orange overlaid with NCS1 and 0.5% DMSO purple, the residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same peak for example Tyr129 and Tyr31, or those which were attenuated by the presence of DMSO including Trp30 and Trp103. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

2.3.3 Introduction to ligand binding studies

With complete resonance assignment of a protein it is then possible to study the interactions with another protein, peptide or ligand. Using NMR spectroscopy to do this provides us with information on any structural changes and indicates the residues and hence, the location involved in any interaction between the two partners. It can also be used to determine binding affinity data; however, this can be complicated by chemical exchange at stoichiometric values of ligand which can cause peak broadening, preventing an accurate peak assignment and thus prevent accurate extrapolation of K_D. Therefore other methods outlined in subsequent sections 2.3.5 and 2.3.6 were favoured.

There are a number of ways to determine the residues implicated in an interaction with a binding partner or peptide of a particular protein. One of the simplest experiments to do would be a 2-D ^{1}H ^{15}N HSQC experiment. An initial spectrum of the protein alone would be used as a control spectrum followed by a secondary spectra of the protein and the ligand or binding partner. To determine the residues involved in any interaction the two spectra can be overlaid, the peaks that shift due to a possible interaction can be tracked and the changes in chemical shift (peak perturbations) calculated using equation 2.1.[268] Where ΔH and ΔN are, respectively, the proton and nitrogen chemical shift changes, and 0.15 is the scaling factor derived from the gyromagnetic ratios of the nitrogen and hydrogen nuclei.[269]

$$\Delta\delta = [\,(\Delta H)^2 + (0.15\,\Delta N)^2\,]^{1/2} \quad 2.1$$

One may also observe attenuation (disappearance) of the peak relating to an amino acid residue; this can be indicative of the protein being in an intermediate exchange state, or could correspond to aggregation of the protein. Although the results from these experiments enable the determination of an interaction if any, between a protein and a ligand or binding partner, a more precise approach uses NOE-based experiments. As NOE is a through space measurement and experimentally allows the identification of hydrogens that are within around 5Å of each other, whether that be within the protein, within the ligand itself or between a protein and a ligand. A ^{13}C ^{15}N filtered NOE experiment is used to selectively observe the interactions between isotopically ^{13}C ^{15}N labelled proteins and unlabelled ligands or small molecules that may be undergoing a binding interaction.

The use of NMR spectroscopy for biophysical assessment of the binding interactions between a protein and ligand, means we are able to elucidate interaction hotspots between the protein and on the ligand itself. It is an extremely useful method to further

our knowledge and understandings on the structure of proteins, enabling the design of small molecules to target them. It can be used successfully on its own, or as in this project it can be used alongside a number of other biophysical techniques to further the structural understanding of the molecules that are important in almost all biological processes.

2.3.4 First Generation Characterisation: Ligand Screening by NMR Spectroscopy

General experimental procedures

Full details of the analysis techniques and experimental procedures, including the conditions used to acquire the screening data of the small molecules with NCS1 by NMR spectroscopy, can be found in Chapter 6.3.2.7. However briefly, for NMR investigation all NMR spectra were collected at 298K, on either a Bruker AVANCE II+ 600 MHz or 800 MHz spectrometer, equipped with 5 mm TCI triple-resonance cryoprobes.

All protein samples were prepared in the NMR buffer (50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$, pH 6.4) and protein samples were from the same 0.5 mM (m/v) NCS1 stock solution, the NMR data collected for investigation included 2-D ^{1}H ^{15}N HSQC and 2-D ^{1}H ^{13}C HSQC TOCSY spectra. The 2-D ^{1}H ^{15}N HSQC binding screens were carried out at a protein: inhibitor ratio of 1:10 whereby the protein concentration was 50 µM and the inhibitors 500 µM respectively. Inhibitor 1 interactions with NCS1 were also investigated at a 1:1 ratio as at the P:L ratio of 1:10 caused severe attenuation of the majority of the amino acid residues and so the data was unsuitable for analysis.

The chemical shift perturbations caused upon an interaction of the small molecules with NCS1 are referenced against the interactions of NCS1 with the same concentration (v/v) of DMSO, which is controlled to be at the same percentage as in the presence of inhibitor and never exceeded 5% (v/v).

Binding data was extrapolated from a 2-D ^{1}H ^{15}N HSQC spectra of 50 µM ^{1}H ^{15}N NCS1 and the chemical shift changes calculated using equation 2.1 (Section 2.3.3).[268] This is known as the “chemical shift perturbation”. The total number of residues perturbed was calculated and from this value the top 10% and 20% ranges were determined. The residues whose chemical shift perturbation sum fell within these two ranges were defined as being within the top 10 and top 20 percentile respectively. Therefore the amino acids in the top 10 and 20 percentile ranges for each individual inhibitor was determined

The changes in chemical shift were then plotted against the protein sequence as a histogram; missing peaks were either attenuated, or un-assigned. This could be due to the shifts being overlaped with other peaks, or where two residues occurred on one peak initially in the original protein assignment. This would, therefore, prevent an accurate transfer of assignment to the spectra of NCS1 in the presence of the ligand. The residues of the top 10 percentile and top 20 percentile were then mapped onto the crystal structure of NCS1. For the purposes of the discussion of the changes in the aliphatic and aromatic methyl spectra, the methyl groups of the amino acid residues are labeled in single letter nomenclature a=alpha, b=beta.

2.3.4.1 Results and Discussion

The first generation of inhibitors included three compounds designed and synthesised through the computational process and synthetic routes that can be found in sections 2.1 and 2.2 respectively (Figure 2.27). As previously mentioned these inhibitors were selected for synthesis and tested against the NCS1 D2R peptide interaction after a rigorous computational pharmacophore screen, followed by systematic computational analysis as described in 2.1. The results of the computational docking predicted binding of the inhibitors within the targeted hydrophobic pocket of NCS1 around the key residues and the predicted poses can be found in 2.1.6.

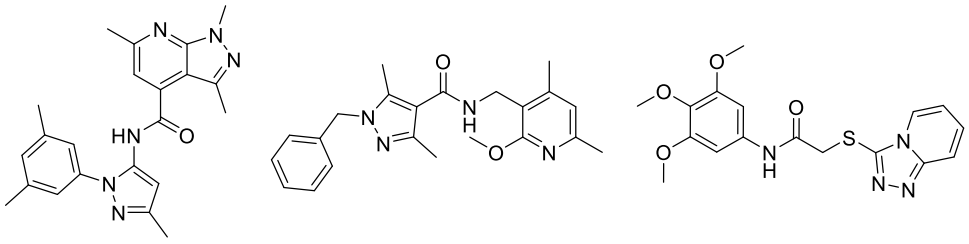


Figure 2.27- Chemical structures of the three inhibitors labelled accordingly, synthesised as described in Section 2.2.

The initial NMR binding screen of the three inhibitors indicated a change in protein chemical shifts when each inhibitor was introduced to NCS1, albeit to differing degrees. Inhibitor 1 at the initial ten-fold ligand protein concentration of 500 μM, caused severe peak broadening and also changes in chemical shift of a number of the amino acid residues (Figure 2.28). Such severe line broadening is indicative of either the protein being in an intermediate exchange regime or aggregation. At this concentration residues were either attenuated or they were ambiguously assigned due to being overlaped with other peaks, therfore preventing an accurate transfer of assignment. Thus the data was not further analysed and so the protein ligand (P:L) ratio was decreased.

When the P:L ratio was decreased to 1: 1, (50 μM) the effect of the inhibitor appeared less severe (Figure 2.29). Peaks belonging to residues of NCS1 such as Ile80, Trp103 and Ile161 were likewise attenuated.

At the 1:1 P:L ratio, there were fewer peak shifts and peak attenuations of NCS1 resonances in comparison to the effects of a higher concentration of ligand. The changes in ^{1}H and ^{15}N chemical shift is a useful indicator of the degree of influence the small molecule has at that particular position on the protein backbone (Section 2.3.3). At 50 μM of Inhibitor 1 the residues within the top 10 percentile (Figure 2.30) experience only very small shift changes (between 0.04 - 0.025); these residues include: Tyr52, Asn70, Tyr129, Gly41, Val180, Phe55, Arg102, Phe34, Lys53, Asn134 and Glu171, with a high propensity to the α2 and α3 regions located within EF-1. The residues of the top 20 percentile included: Thr92, Gln49, Asp44, Gly172, Leu189, Ser83, Gly59, Glu141, Asp73, Asn143, Glu142, there appears to be no specificity as those residues affected are found across the four EF hands.

At 50 μM Inhibitor 1, the number of residues who were attenuated was significatly reduced. Those residues who experienced a change in chemical shift was not only greater, their position was infact on the correct face of the protein and closer to the targeted hydrophobic pocket (Figure 2.31).

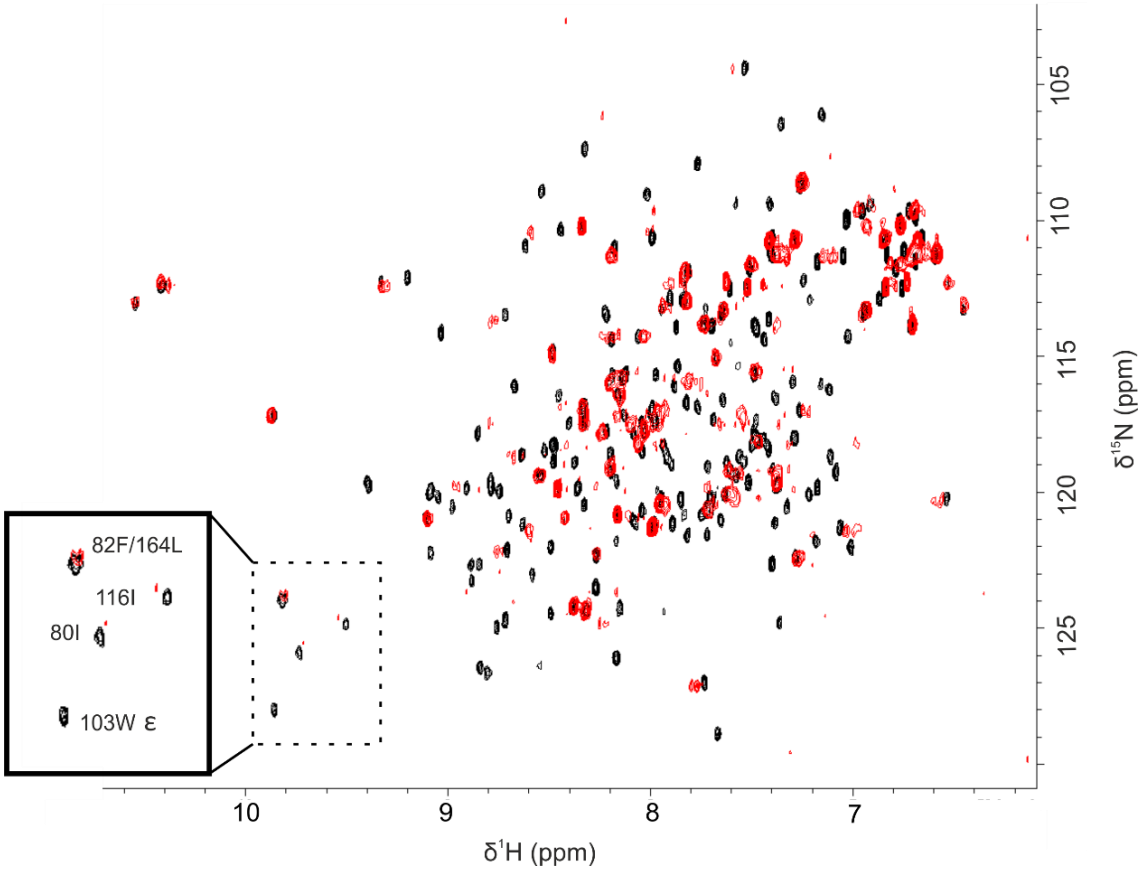


Figure 2.28- Ligand binding screen NCS1 and Inhibitor 1. An overlay of two 2-D ^{1}H ^{15}N HSQC spectra, the control spectra ^{1}H ^{15}N NCS1 NCS1 at a concentration of 50 µM in the presence of 1% DMSO (black). The spectra of ^{1}H ^{15}N NCS1 50 µM with Inhibitor 1, 500 µM in the presence of 1% DMSO (P:L 1:10) (red), show severe peak attenuation. The residues highlighted within the box were selected to indicate changes in specific residues. ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer 50 mM tris HCl, 50 mM NaCl 5 mM $CaCl_2$ pH 6.4, acquired on Bruker AVANCE II+ 600 MHz spectrometer at 298K.

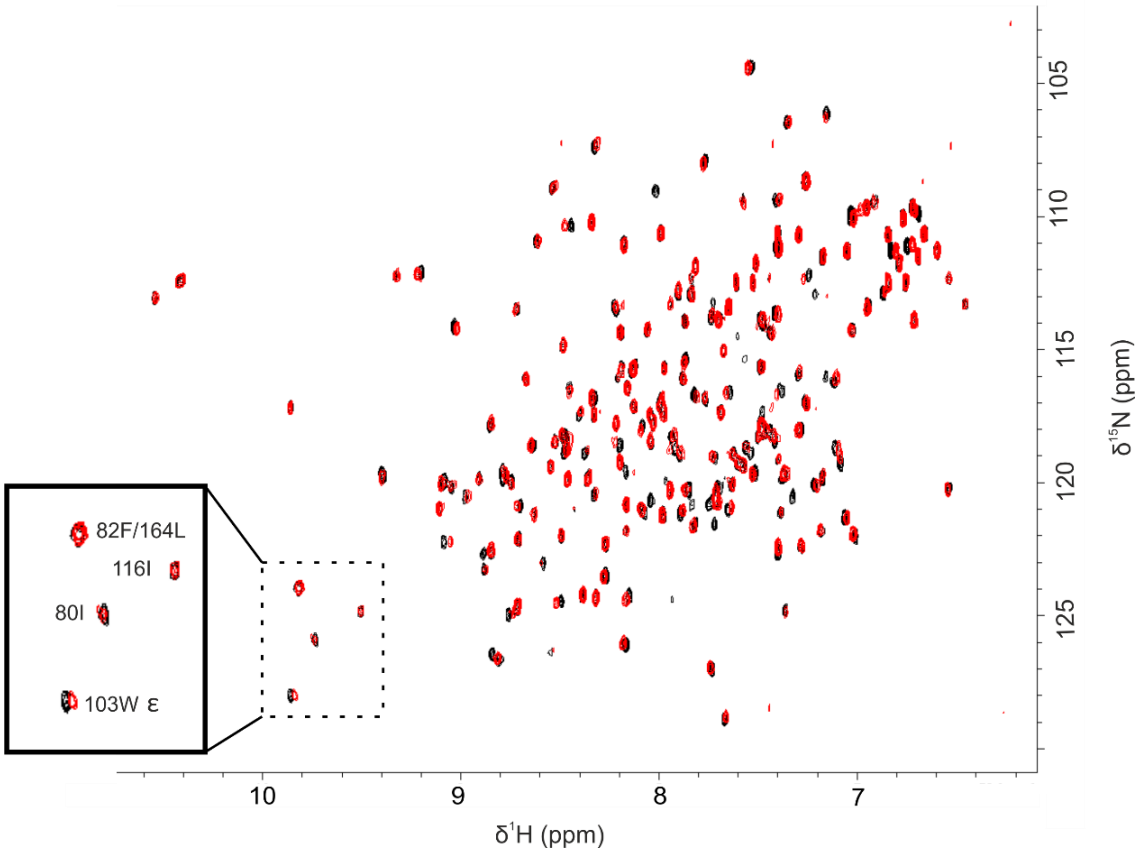


Figure 2.29- Ligand binding screen NCS1 and Inhibitor 1. An overlay of two 2-D ^{1}H ^{15}N HSQC spectra, the control spectra ^{1}H ^{15}N NCS1 at a concentration of 50 µM in the presence of 0.1% DMSO (black). The spectra of ^{1}H ^{15}N NCS1 50 µM with Inhibitor 1 50 µM in the presence of 0.1% DMSO (P:L 1:1) (red), showing less severe peak attenuation and some changes in the chemical shift of a number of residues. The residues highlighted within the box were selected to indicate changes in specific residues. ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer 50 mM tris HCl, 50 mM NaCl 5 mM $CaCl_2$ pH 6.4, acquired on Bruker AVANCE II+ 600 MHz spectrometer at 298K.

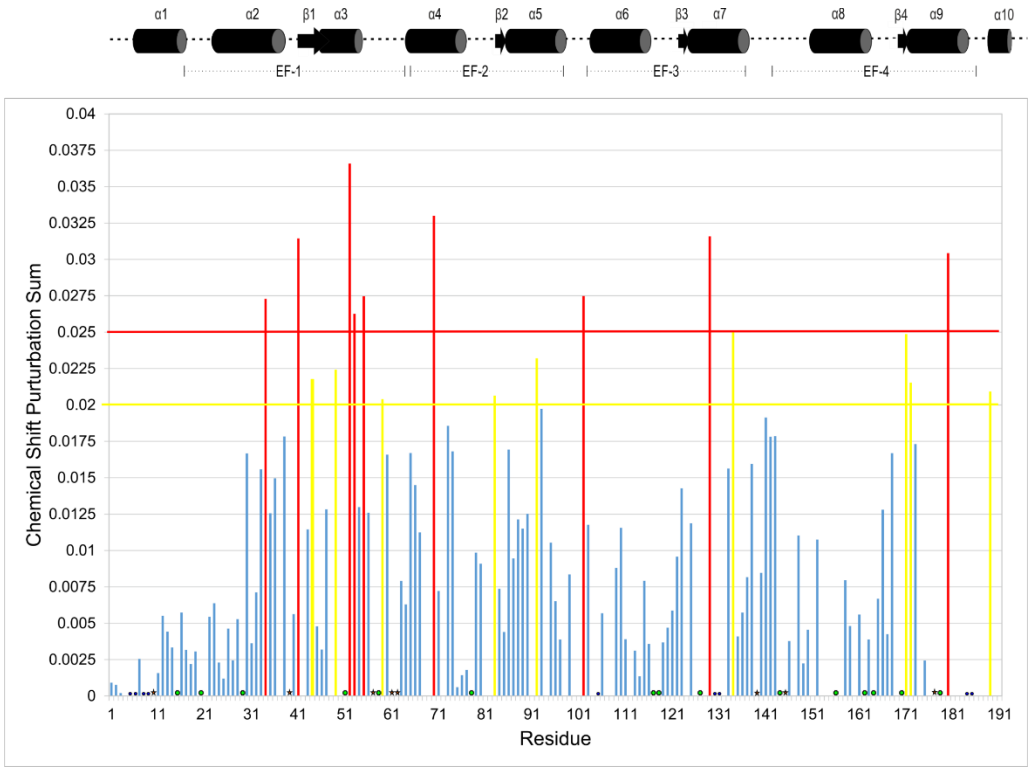


Figure 2.30– Inhibitor 1 Peak perturbation Histogram. A Histogram of the perturbations of NCS1 residues from ligand binding screen NCS1: Inhibitor 1, 50 μM. Those residues whos chemical shift underwent a change in ppm in the top 10% or over can be seen above the red line and those residues include: Tyr52, Asn70, Tyr129, Gly41, Val180, Phe55, Arg102, Phe34, Lys53, Asn134 and Glu171. The yellow line represents those residues who's change in chemical shift was within the top 20% of changes and includes: Thr92, Gln49, Asp44, Gly172, Leu189, Ser83, Gly59, Glu141, Asp73, Asn143 and Glu142. • Unassigned residues ✿ Single peaks assigned with two residues ⋆ Proline.

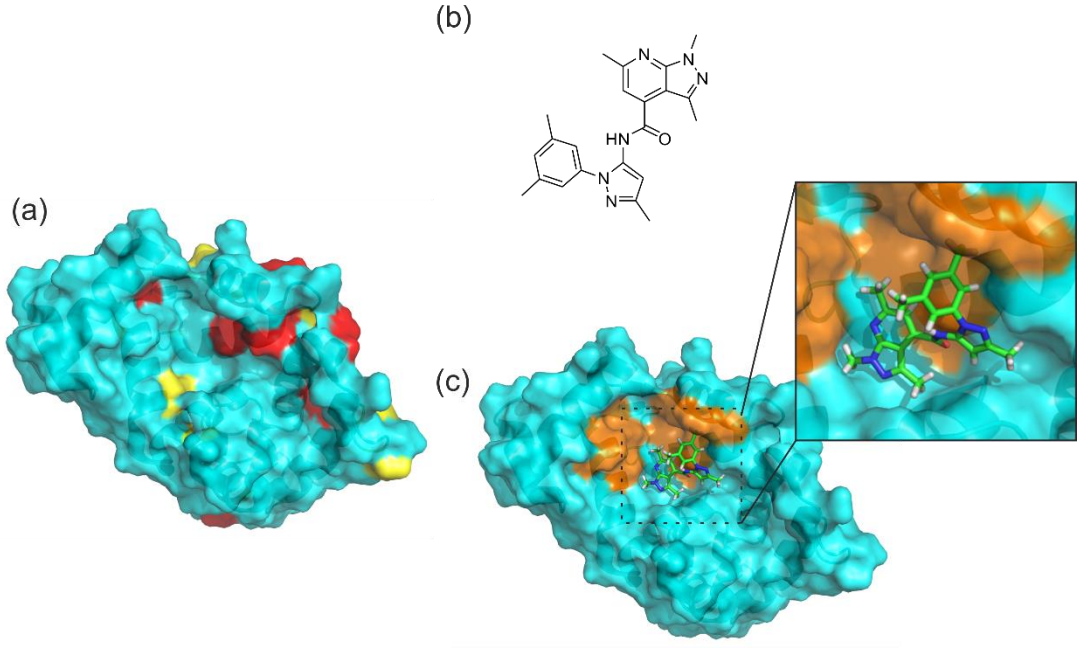


Figure 2.31– (a) Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 1 at 50 μM determined from the corresponding histogram top 10% coloured in red and top 20% in yellow (PDB 5AER). **(b)** Chemical structure of Inhibitor 1. **(c)** Prediction of the binding pose for the interaction between Inhibitor 1 and NCS1 (Section 2.1.6) those key hotspot residues highlighted in orange.

As well as the change in chemical shift of the backbone residues brought about through the effects of Inhibitor 1, it is also useful to evaluate any changes in the sidechains. A complete analysis of the changes in sidechain is hampered by the fact that only 42.7% of the sidechains is assigned. Therefore, a more superficial analysis was conducted to determine which resonances are affected by the presence of the ligand. In the case of the sidechains containing methyl groups, this may appear as complete attenuation of the relative peaks, or partial attenuation in those cases where the residue contains two methyl groups such as leucine but only one is attenuated.

With the sidechain investigations of Inhibitor 1 the difference observed at a protein to ligand ratio of 1:10 was not as severe as that of the sensetive amide backbone experiments. Therefore for equivalent comparison between all first generation inhibitors, the difference between the control ^{1}H ^{13}C HSQC spectra of 50 μM ^{1}H ^{15}N ^{13}C NCS1, 1% DMSO with the ^{1}H ^{13}C HSQC spectra of 50 μM ^{1}H ^{15}N ^{13}C NCS1 and Inhibitor 1 500 μM (1:10 P:L ratio Figure 2.32) was used to investigate the interactions. The investigations indicate an effect which is found to span across all four EF hand domains of NCS1, further details on the secondary structure of NCS1 can be found in the introduction (Section 1.1, Figure 1.5).

Of the twenty two methyl-containing sidechain residues affected, ten were completely attenuated which indicates the possibility that methyl sidechains that were once exposed, have become buried in the presence of Inhibitor 1. These residues include; Leu43, Ile51, Val68, Thr92, Leu101, Ala104, Ile124, Leu138, Ile152 and Ala182, they are located across the four EF hands and are present in α helix, β sheet and loop regions. Those 12 residues that underwent a partial attention of a single methyl group include Leu_a16, Val_a27, Val_b91, Leu_b97, Leu_b107, Leu_b110, Leu_b125, Ile_a128, Val_a132, Leu_b164, Ile_a179 and Val_a180, with a high propensity towards the EF-3 region, however some are also found on EF-1, EF-2 and EF-4.

Sidechain assignment is not limited to those methyl-containing residues, aromatic residues such as phenylalanie, tyrosine and tryptophans were also assigned, the effects of Inhibitor 1 on these residues in comparison to the DMSO containing spectra is severe (Figure 2.33). Here we observe the attenuation of a number of residues including Phe22, Tyr52, Phe67, Phe85, Tyr108, Tyr129, Phe153 and Phe169, located on EF-1, EF-2, EF-3 and EF-4.

This methyl and aromatic sidechain data is in concordance with the 2-D ^{1}H ^{15}N HSQC data collected at the same protein to ligand ratio, indicating the widespread affect of Inhibitor 1 on the residues of NCS1 at a high concentration.

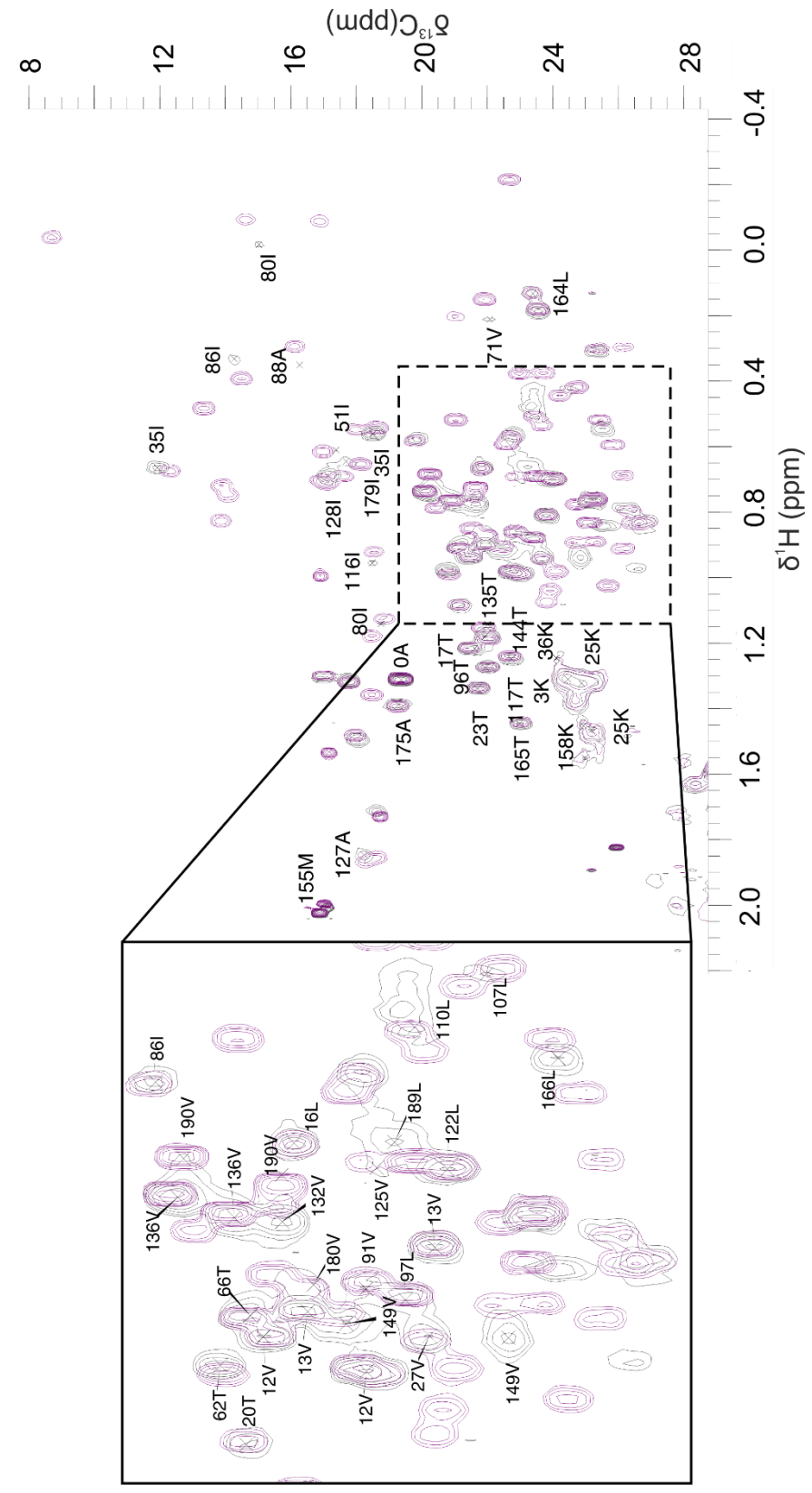


Figure 2.32- An overlay of two methyl sidechain regions of NCS1 with and without Inhibitor 1. The ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1% DMSO (purple) overlaid with the ^{1}H ^{13}C TOCSY spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 1 500 µM with 1% DMSO (grey). Those residues of NCS1 related to the Inhibitor 1 spectra for which the transfer of assignment was possible can be seen, there a number of notable attenuations including Leu89, Ile116 and Ile124. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

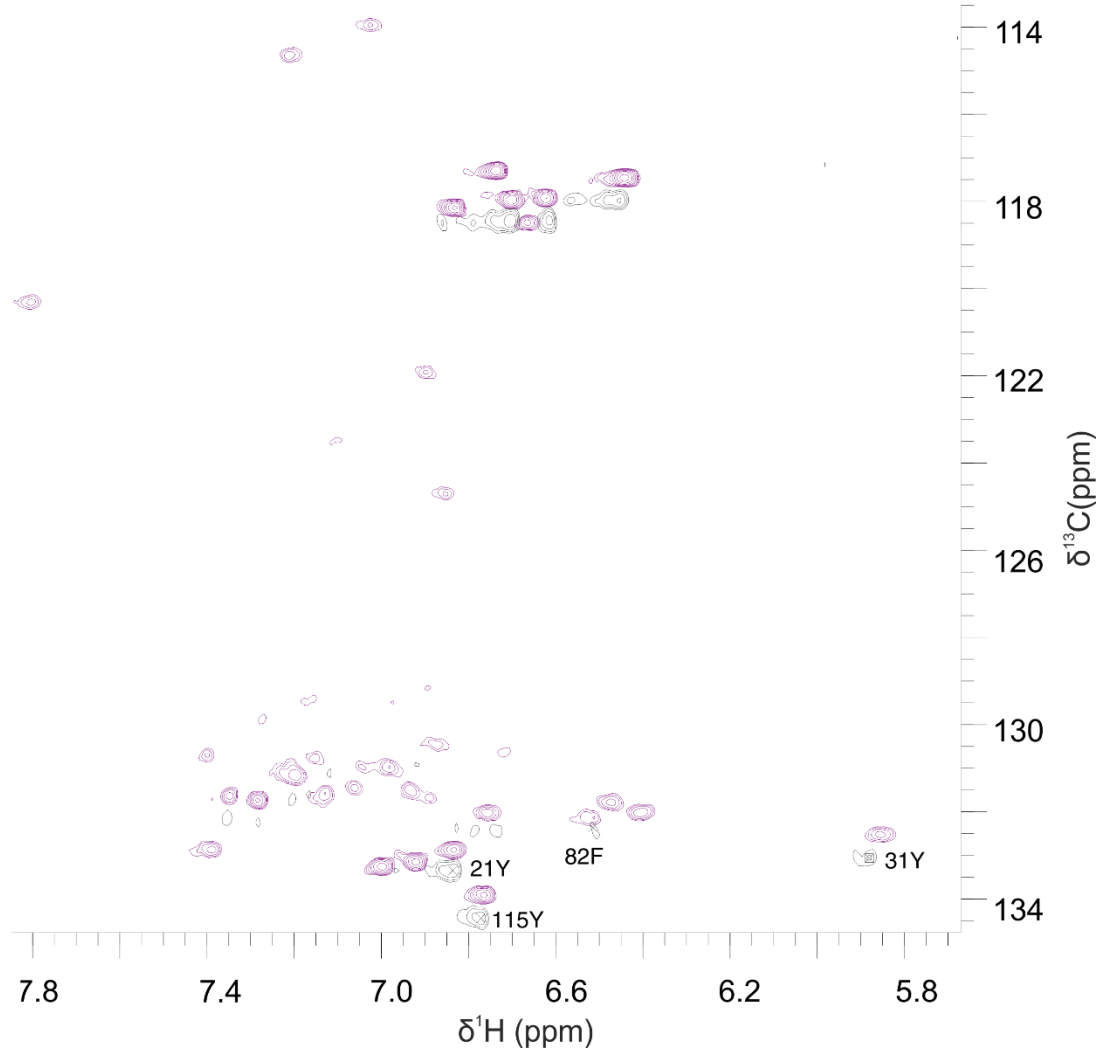


Figure 2.33- Two overlaid aromatic sidechain spectra of NCS1 with and without Inhibitor 1. The aromatic ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1% DMSO purple overlaid with the ^{1}H ^{13}C spectra of^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 1 500 µM with 1% DMSO, the residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same peak for example or those which were attenuated by the presence of Inhibitor 1 including; Tyr52, Phe22, Phe67, Phe85, Tyr108, Tyr129, Phe153 and Phe169. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

As with Inhibitor 1, the first NMR spectroscopy experiment with Inhibitor 2 was carried out at a P:L ratio of 1:10, with the results being different to that of Inhibitor 1. The amino acid residues of the protein did not appear to be as strongly effected, in that the number of residues attenuated was reduced. However, those residues that were affected underwent a substantial change in chemical shift. Such residues include Ile80 and Trp103 (Figure 2.34). These changes in chemical shift could suggest a binding interaction occurring between Inhibitor 2 and NCS1 which was further substantiated by intrinsic tryptophan fluorescence (2.3.8).

The chemical shift changes for the effect of 500 μM Inhibitor 2 with NCS1 was calculated as a combined Δ H and Δ N shift. The results of which can be found (Figure 2.35), the residues of the top 10 percentile experienced a change in chemical shift between 0.07 - 0.18. Those residues which undergo changes within this range include: Gly172, Asp44, Lys36, Gln140, Leu166, Gly33, Thr96, located on the EF hands EF-1 and EF-4. Those residues within the top 20 percentile include : Ala65, Thr165, Asp109, Val71, Tyr31, Ile80, Thr23 which are also located EF-2 and EF-3, more specifically α4, α5 and α6.

From this data it is apparent that there is a noteable increase in the chemical shift change, between the NCS1 reference spectrum and the effect of 500 μM Inhibitor 2. When those peaks affected by Inhibitor 2, were mapped onto the crystal structure of NCS1 those residues that were identified as in the top 10 percentile were found to be on the same face as the hydrophobic groove. Albeit not all were in the desired residue range, as they were located primarily on EF-1, EF-2 and EF-3. The results from the residues of the top 20 percentile were less specific as many were on the opposite face of the protein.

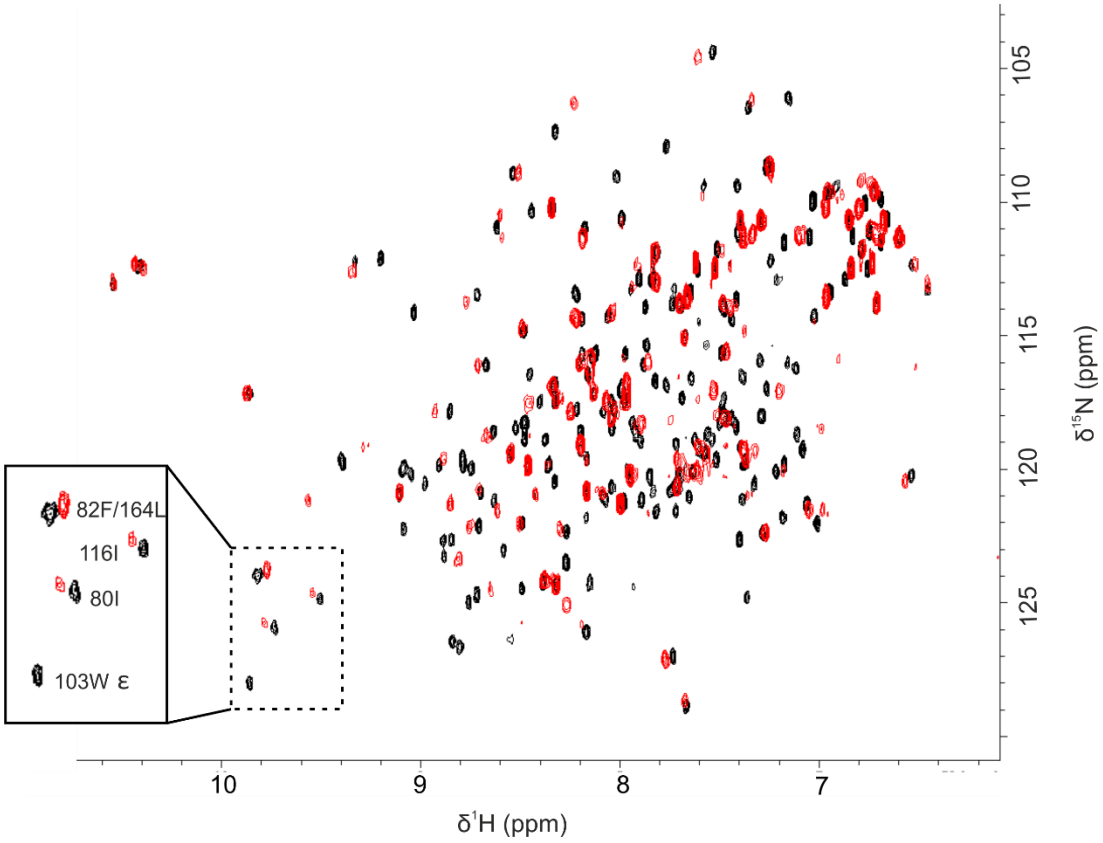


Figure 2.34- Ligand binding screen with Inhibitor 2. An overlay of two 2-D ^{1}H ^{15}N HSQC spectra, the control spectra ^{1}H ^{15}N NCS1 at a concentration of 50 μM in the presence of 1% DMSO (black). The spectra of ^{1}H ^{15}N NCS1 50 μM with Inhibitor 2, 500 μM in the presence of 1.0% DMSO (P:L 1:10) (red) showing attenuation of some peaks and changes in chemical shift indicative of an interaction between the inhibitor and the protein. The residues highlighted within the box were selected to indicate changes in specific residues within a conserved region of the HSQC spectra. ^{1}H ^{15}N NCS1 50 μM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 600 MHz spectrometer at 298K.

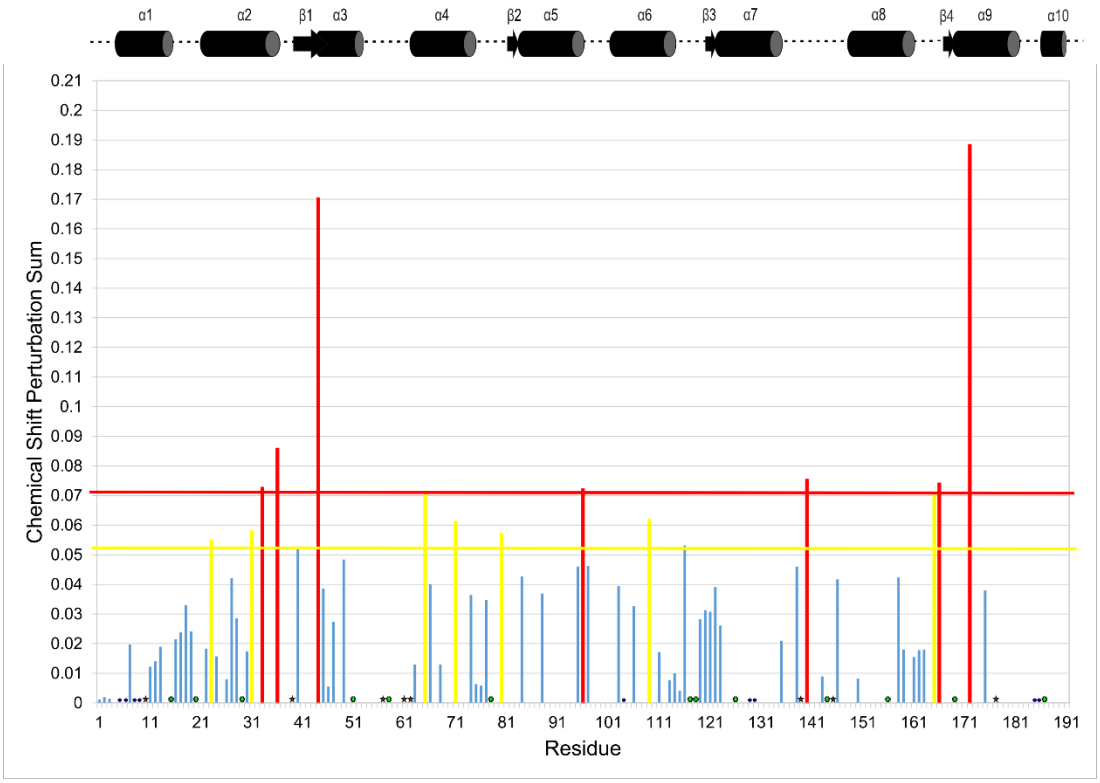


Figure 2.35- Inhibitor 2 Peak Perturbation Histogram. A histogram of the perturbations in NCS1 residues from ligand binding screen NCS1: Inhibitor 2 500 μM. Those residues whos chemical shift underwent a change in ppm in the top 10% or over can be seen above the red line and those residues include: Gly172, Asp44, Lys36, Gln140, Leu166, Gly33, Thr96. The yellow line represents those residues who's change in chemical shift was within the top 20% of changes and includes: Ala65, Thr165, Asp109, Val71, Tyr31, Ile80, Thr23. ● Unassigned residues ✿ Single peaks assigned with two residues ★ Proline.

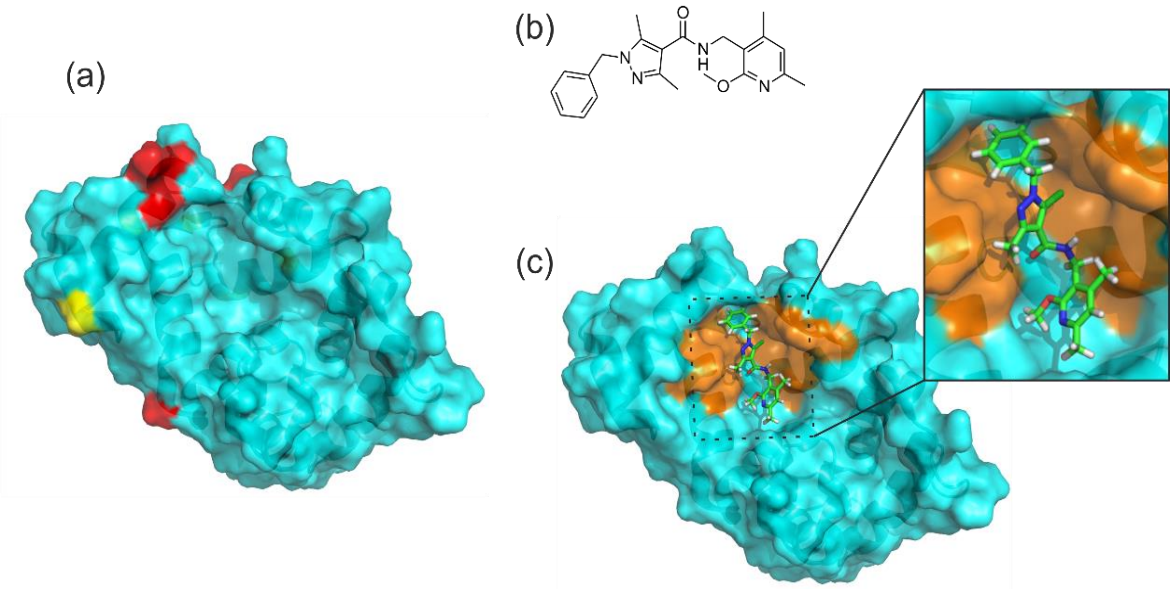


Figure 2.36- **(a)** Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 2 at 1 mM determined from the corresponding histogram top 10% coloured in red and top 20% in yellow (PDB 5AER). **(b)** Chemical structure of Inhibitor 2 **(c)** Prediction of the binding pose for the interaction between Inhibitor 2 and NCS1 (Section 2.1.6) those key hotspot residues highlighted in orange.

As with Inhibitor 1, the possible binding interactions of Inhibitor 2 were not only investigated through determination of changes in chemical shift of the backbone residues. The interactions of the methyl and aromatic sidechains of the residues were also evaluated. This was carried out as before with Inhibitor 1 through the observation of changes in residues between the control spectra and the ligand containing sample (Figure 2.37).

In the case of Inhibitor 2, of the 18 methyl-containing sidechains for which assignment transfer was possible, ten of the residues were completely affected in that all methyl groups were attenuated and they include; Leu43, Ile51, Val68, Thr92, Ala104, Ala104, Val125, Ile179, Ala182 and Val190. These residues are located on α helices within EF-1, EF-2 and EF-3 with the final three residues being clustered in the terminal loop region.

The eight residues that were only partially attenuated include: Leu_a16, Val_a27, Leu_b125, Ile_b12, Ile_a128, Ile_a152, Leu_a164 and Val_b180, located on EF-1, EF-3 and EF-4, unlike the fully attenuated residues there are no partial attenuations on EF-2.

Evaluation of the effect that Inhibitor 2 has on the aromatic sidechain residues, indicate that Inhibitor 2 has a greater effect on the tyrosine (Tyr) ε residues that are located around 118 ppm (Figure 2.38). In comparison to the DMSO control (Figure 2.26 b), two of these such residues Tyr ε 52 and Tyr ε 108 are attenuated in the presence of Inhibitor 2. The remaining aromatic sidechain residues affected by Inhibitor 2 include Phe22, Phe67, Phe85, Phe153 and Phe169, with a high propensity towards α helical locations across the four EF hands.

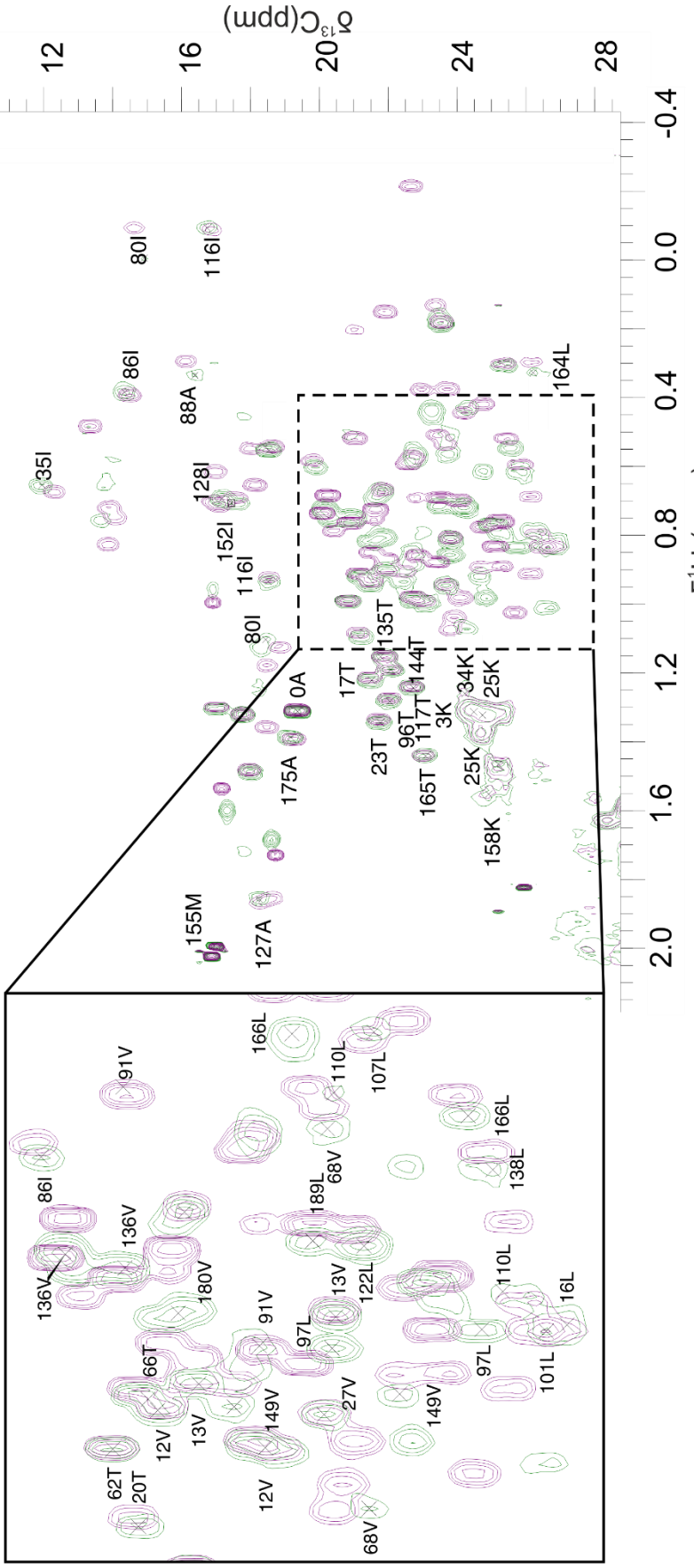


Figure 2.37- An overlay of two methyl sidechain regions of NCS1 with and without Inhibitor 2. The ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1% DMSO (purple) overlaid with the ^{1}H ^{13}C TOCSY spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 2 500 µM with 1% DMSO (green). Those residues of NCS1 related to the Inhibitor 2 spectra for which the transfer of assignment was possible can be seen, there a number of attenuations including Leu89. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

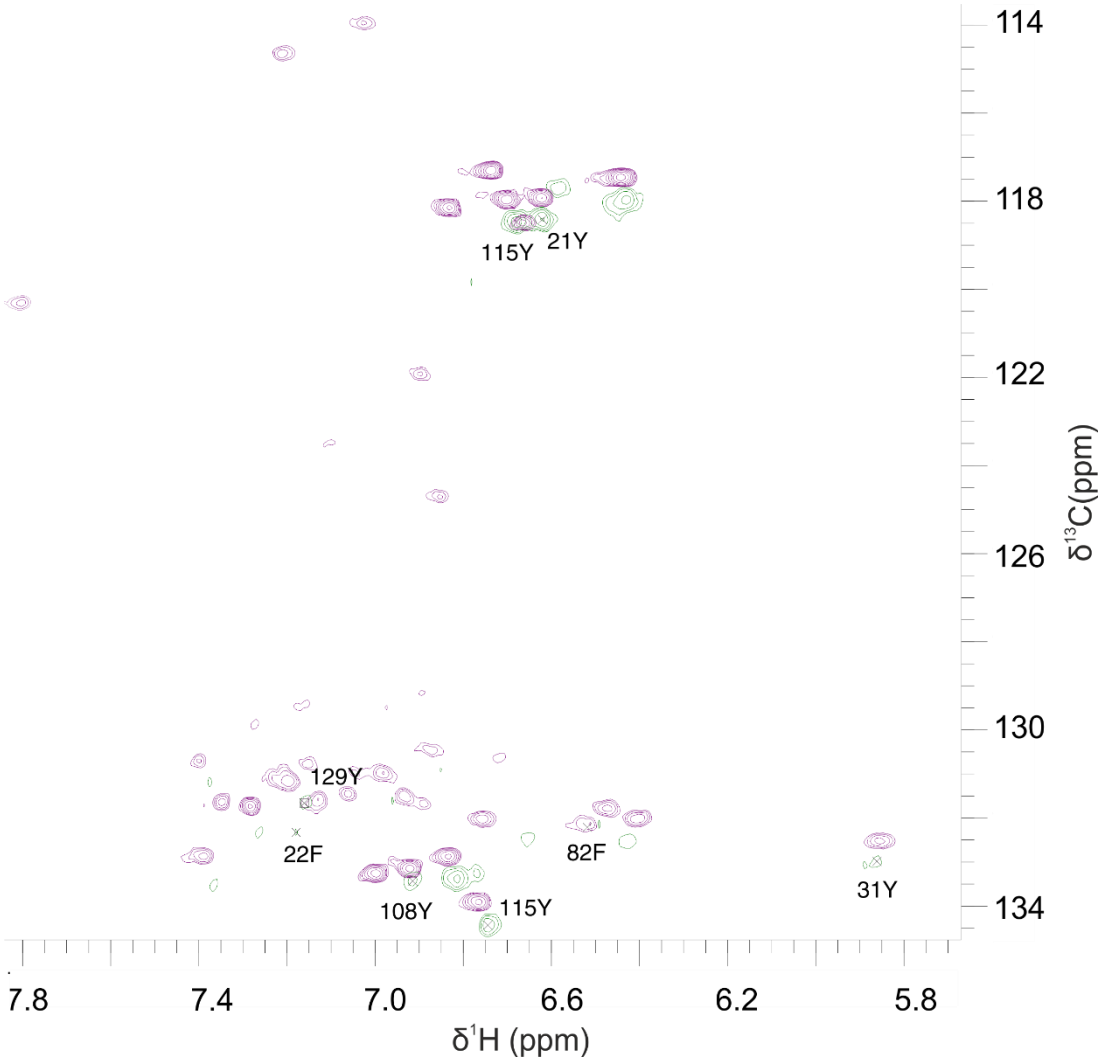


Figure 2.38- Two overlaid aromatic sidechain spectra of NCS1 with and without Inhibitor 2. The aromatic ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1% DMSO purple overlaid with the ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 2 500 µM with 1% DMSO (green), the residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same resonance or those which were attenuated by the presence of Inhibitor 2 including; Phe22, Tyr52, Phe67, Phe85, Tyr108, Phe153 and Phe169. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

Inhibitor 3 was also subjected to the same spectroscopic experiments as inhibitors1 and 2. At the same initial protein: ligand ratio of 1: 10 the inhibitor appears to not have the same effect on the protein as the first two inhibitors. This can be seen in in the 2-D ^{1}H ^{15}N HSQC overlay (Figure 2.39).

Although Trp103 appears to undergo a change in chemical shift and Ile80 appears slightly attenuated, Ile161, Phe82 and Leu164 were unaffected. This data is in concordance with that of the chemical shift perturbations (Figure 2.40), in that the residues of the top 10 percentile, undego changes in chemical shift over a smaller range than both that of Inhibitors1 and 2. Residues within this range include: Leu183, Val180, Asp44, Gln181, Lys174, Tyr108, Arg102, Glu171, Lys53, Gly41, Lys106, Lys100, located in three clusters within the protein. The largest cluster of residues peturbed are found on the α helices within EF-4 however, the remaining two clusters of residues peturbed can also be found along α2, α5 and α6.

Those residues within the top 20 percentile include: Phe34, Glu141, Ala65, Tyr129, Asn70, Leu107, Cys38, Gln172, Gly33, Phe48, Ile152, Ser173. The subsequent mapping of these residues onto the crystal structure appear to indicate the location to be mainly around the same face as the hydrophobic pocket, with a high propensity to the helices contained within EF-1.

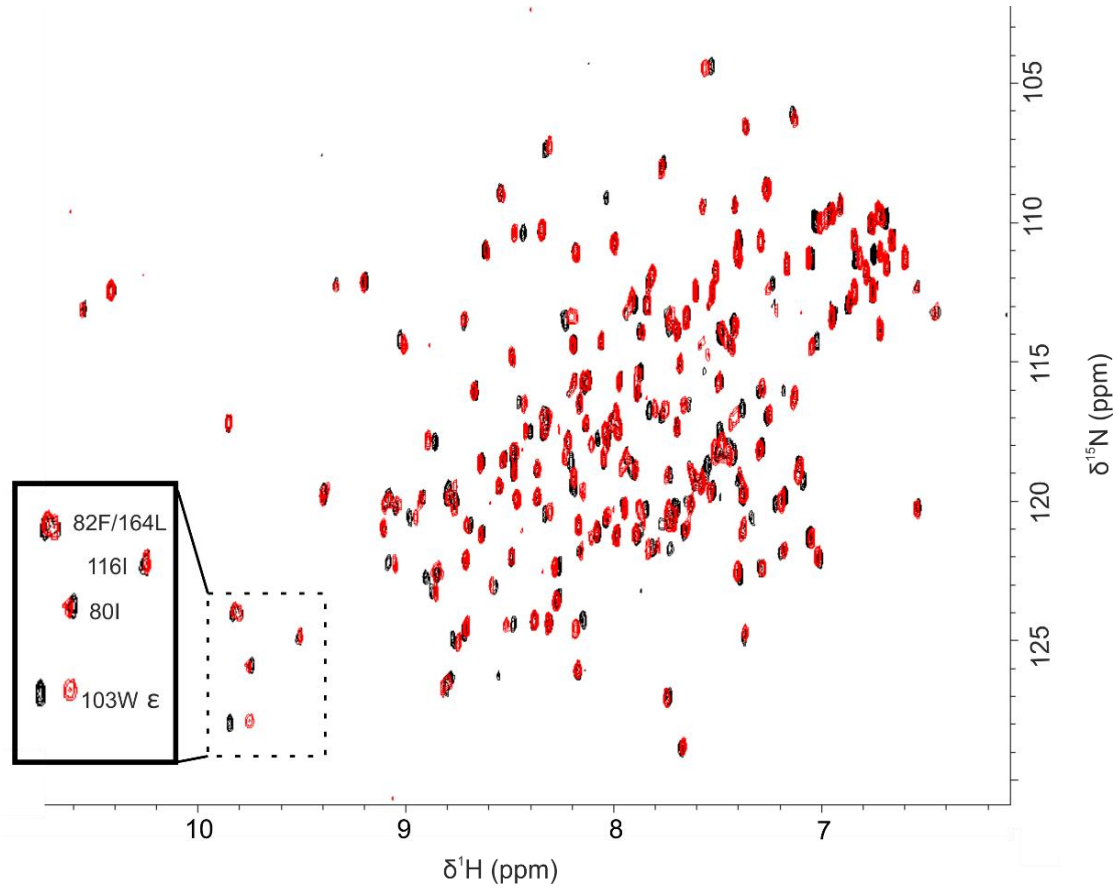


Figure 2.39- Ligand binding screen with Inhibitor 3. An overlay of two 2-D ^{1}H ^{15}N HSQC spectra, the control spectra ^{1}H ^{15}N NCS1 at a concentration of 50 µM in the presence of 1.0% DMSO (black). The spectra of ^{1}H ^{15}N NCS1 50 µM with Inhibitor 2 500 µM in the presence of 1.0% DMSO (P:L 1:10) (red) showing attenuation of some peaks and changes in chemical shift indicative of an interaction between the inhibitor and the protein, however the quantity of shifts perturbed could possibly indicate non-specific binding. The residues highlighted within the box were selected to indicate changes in specific residues within a conserved region of the HSQC spectra. ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 600 MHz spectrometer at 298K.

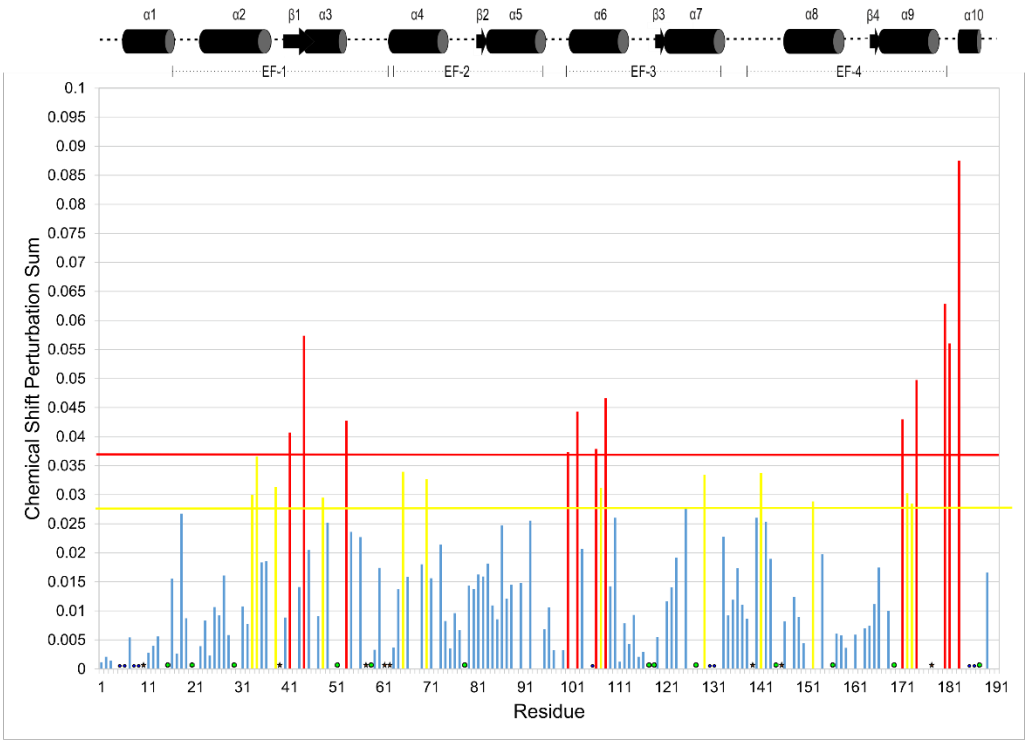


Figure 2.40– Histogram of the Perturbations in NCS1 residues from ligand binding screen NCS1: Inhibitor 3 500 μM**.** Those residues whos chemical shift underwent a change in ppm in the top 10% or over can be seen above the red line and those residues include: Leu183, Val180, Asp44, Gln181, Lys174, Tyr108, Arg102, Glu171, Lys53, Gly41, Lys106, Lys100. The yellow line represents those residues who's change in chemical shift was within the top 20% of changes and includes: Phe34, Glu141, Ala65, Tyr129, Asn70, Leu107, Cys38, Gln172, Gly33, Phe48, Ile152, Ser173. ● Unassigned residues ✿ Single peaks assigned with two residues ★ Proline.

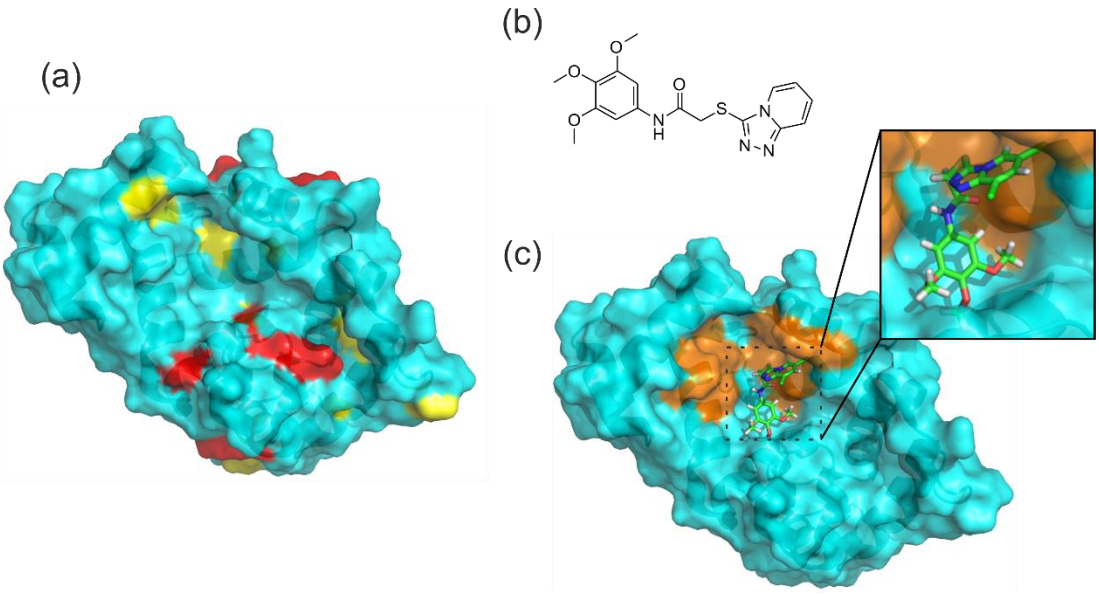


Figure 2.41- (a) Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 3 at 1 mM determined from the corresponding histogram top 10% coloured in red and top 20% in yellow (PDB 5AER). **(b)** Chemical structure of Inhibitor 3. **(c)** Prediction of the binding pose for the interaction between Inhibitor 2 and NCS1 (Section 2.1.6) those key residues indicated by Ames *et al.*highlighted in orange.

Again as with the previous two inhibitors, the possible binding interactions of Inhibitor 3 with the backbone residues of NCS1 were investigated alongside the interactions of the methyl and aromatic sidechains of the residues. To do this the difference between the control ^{1}H ^{13}C HSQC TOCSY spectra of 50 µM ^{1}H ^{15}N ^{13}C NCS1 and 1% DMSO with the ^{1}H ^{13}C HSQC TOCSY spectra of 50 µM ^{1}H ^{15}N ^{13}C NCS1 and Inhibitor 3 500 µM (Figure 2.42), at a 1:10 P:L ratio were evaluated.

In comparison to the results of Inhibitor 2 at the same concentration, the effects of Inhibitor 3 on the sidechain methyl groups are substantially reduced. None of the methyl-containing residues are fully attenuated and only six residues undergo partial attenuation induced through the effects of the inhibitor. The partial attenuations occur over the four EF hands, they include the residues Leu_a16, Ile_b51, Val_a71, Ile_b128, Ile_a152 and Val_b190, present on α helices and the terminal loop.

The Tyr ε residues are un-attenuated in the spectra of Inhibitor 3 (Figure 2.43) and the aromatic sidechain residues that underwent an interaction with Inhibitor 2 were also affected by Inhibitor 3 and they include; Phe22, Phe67, Phe85, Phe153 and Phe169.

The effects of Inhibitor 2 and Inhibitor 3 on the methyl and aromatic sidechain residues of NCS1 are in concordance with the collected 2-D ^{1}H ^{15}N HSQC spectra, from which the chemical shift perturbation sum of the backbone residues was calculated. In both cases we observe a more subtle change in the NCS1 backbone and sidechain residues in the presence of Inhibitor 3, with a greater effect being incurred in the presence of Inhibitor 2. The results of the initial NMR spectrosopy binding screens for all three Inhibitors, present promising data that indicates all three inhibitors are interacting with the protein, allbeit at different degrees.

The somewhat non specific binding, away from the intended targeted residues could merely be concentration dependent , in that too high a ligand concentration induces non specific binding. In the case of Inhibitor 1, too high a concentration was found to possibly have a deleterious affect on the protein, observed through the severe peak broadening and attenuation of a large number of residues across the protein. However these effects could also indicate that a lower concentration of Inhibitor 2 and Inhibitor 3 are unable to saturate NCS1, this could be indicative of a weak interaction between the ligand and the protein. Therefore to investigate this further and determine the binding affinity, the three inhibitors were subsequently investigated using Isothermal Titration Calorimetry (ITC) 2.3.5.

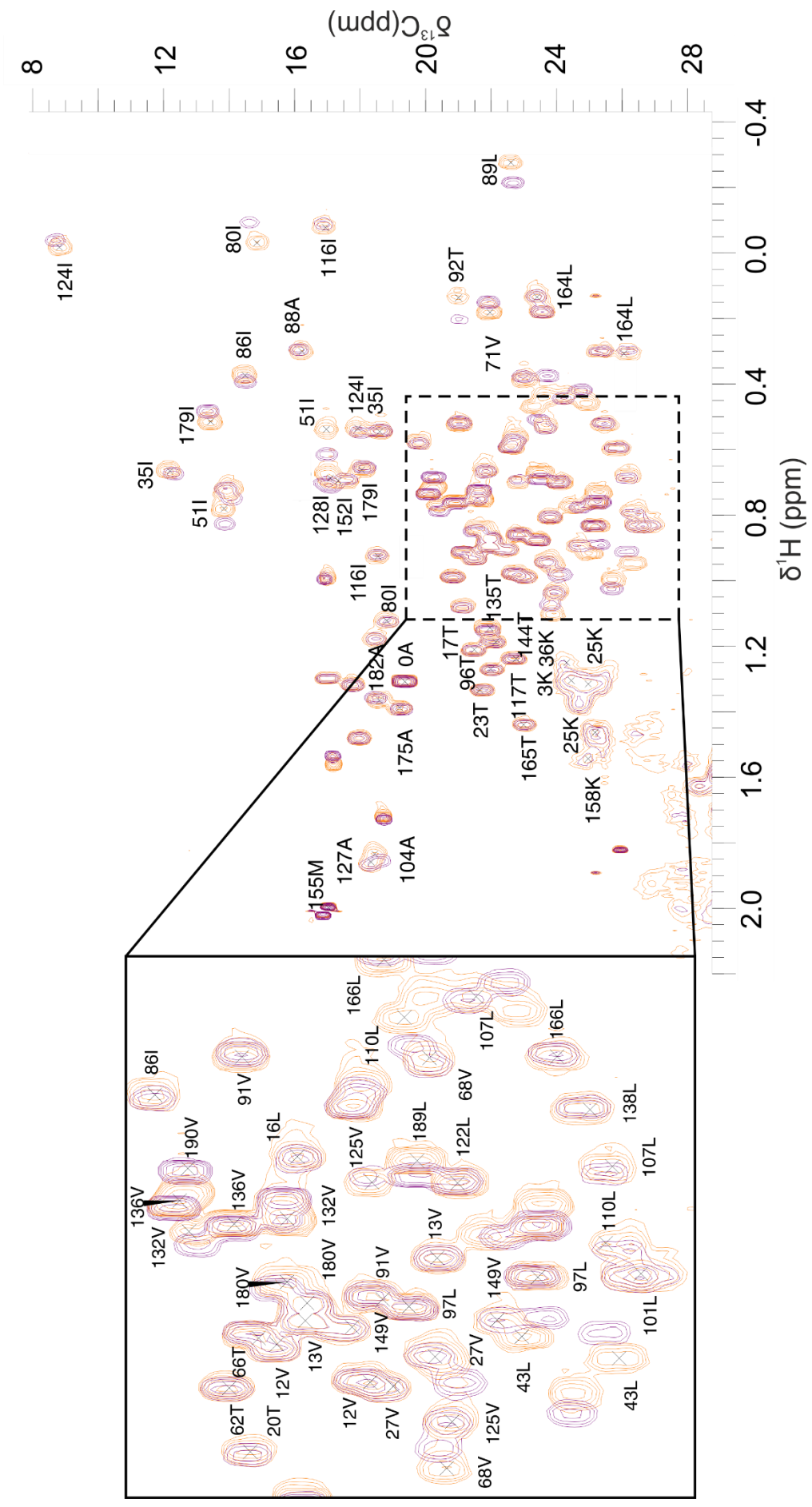


Figure 2.42- An overlay of two methyl sidechain regions of NCS1 with and without Inhibitor 3. The ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1.0% DMSO (purple) overlaid with the ^{1}H ^{13}C TOCSY spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 3 500 µM with 1% (orange). Those residues of NCS1 related to the Inhibitor 3 spectra for which the transfer of assignment was possible can be seen, there a number of notable changes in chemical shift including Ile80, Leu89 and Ile124. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

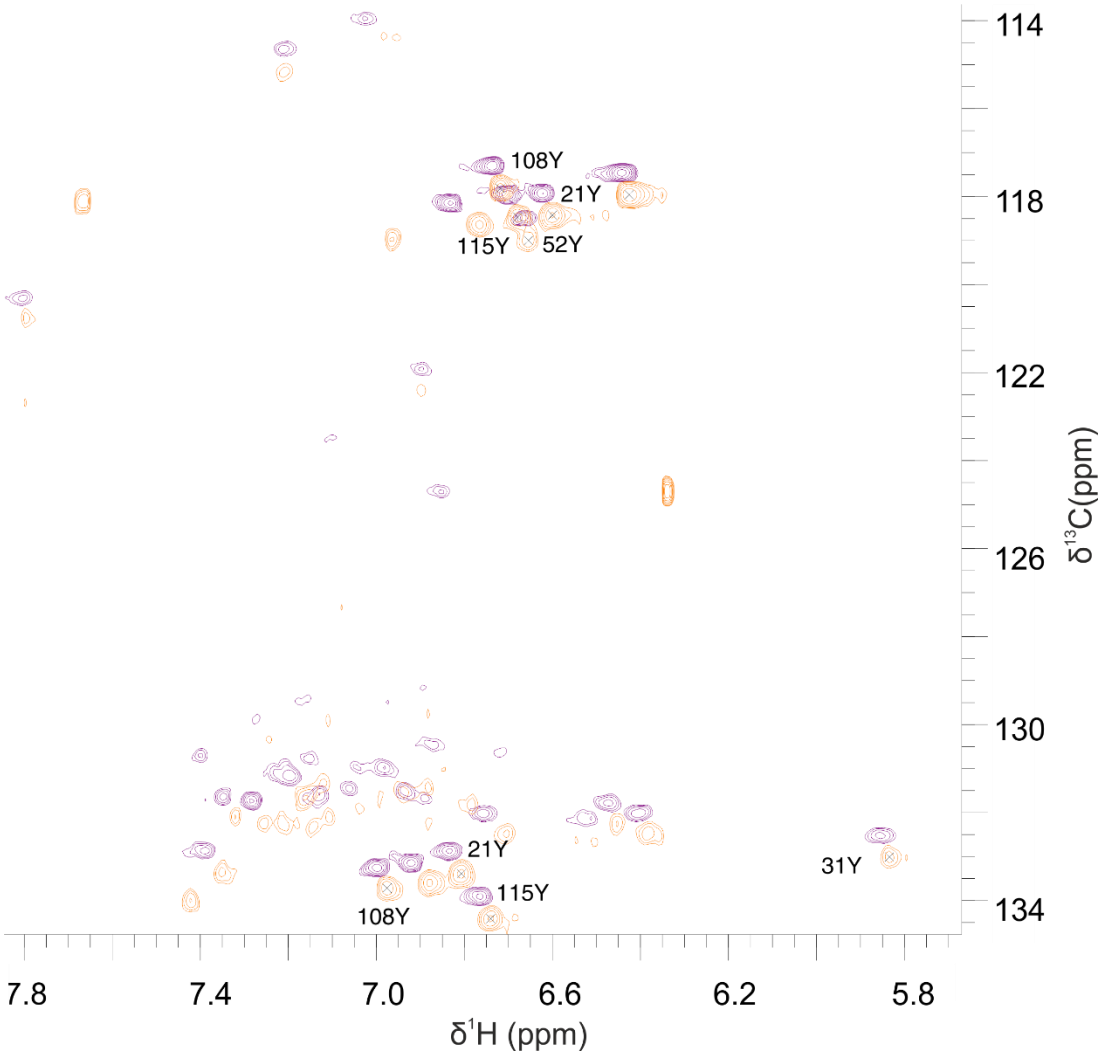


Figure 2.43- Two overlaid Aromatic sidechain spectra of NCS1 with and without Inhibitor 3. The aromatic ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 1.0% DMSO purple overlaid with the ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 3 500 µM with 1% DMSO (orange), the residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same peak or those which were attenuated by the presence of Inhibitor 3 including; Phe22, Phe67, Phe85, Tyr129, Phe153 and Phe169F. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

.

2.3.5 Introduction to Isothermal Titration Calorimetry (ITC)

As introduced previously Chapter 1.6.2, isothermal calorimetry is the measure of a change in heat of a physical process, with the sample of interest and reference sample kept at the same constant temperature during the experiment, hence its term isothermal.[210] ITC is able to provide a detailed thermodynamic characterisation of the interaction between a target receptor/ protein and a ligand or secondary binding partner. Such thermodynamic properties include the directly obtained enthalpy (ΔH) and entropy (TΔS) and the derived Gibbs free energy (ΔG) and binding affinity (K_D).[211] The ΔH and TΔS contribute to ΔG, which is directly related to K_D through equations 2.2 and 2.3.

$$\Delta G = \Delta H - T\Delta S \qquad 2.2$$

$$\Delta G = -RTlnK\mathrm{D} \qquad 2.3$$

The ITC calorimeter is composed of two cells, a sample cell into which the titration is performed and a reference cell which contains water.[270] Both cells are maintained at a constant temperature by a heater, which is in turn controlled by a power feedback system.[270]

In the case of an exothermic experiment, upon the addition of the ligands heat is released if there is an interaction between the ligand and the therapeutic target.[270] If both cells are to remain at a constant temperature, the feedback system must reduce the power supplied to the heater of the sample cell, whilst maintaining that supplied to the reference cell.[270] This difference in power between the two cells is recorded as a function of time and can be observed in curve (a Figure 2.44).[270] Subsequent integration of the power curve gives the heat change associated with each addition of ligand to the target in the sample cell (b Figure 2.44).[270]

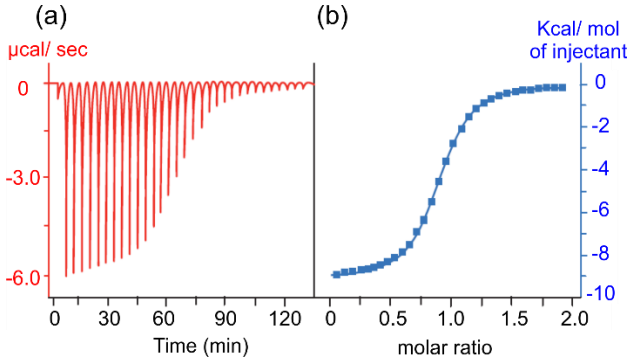


Figure 2.44- Example ITC data. **(a)** Raw data of the difference in power supplied to the heaters as a function of time. **(b)** Integration of heat change per injection with best fitting curve applied (adapted from [270]).

The large heat signals released upon the addition of ligand at the start of the titration, reflect the formation of ligand and target complexes at each injection point.[270] At

the end of the titration, all binding sites on the target receptor should be occupied, the subsequent heat changes occur due to the dilution effects of the ligand in the sample cell.[270]

This can be overcome by performing experimental standards, such as titrating ligand into buffer within the sample cell and titrating buffer into the target/ receptor within the sample cell. The experimental standards can then be subtracted from the experimental data, providing a more accurate representation of the thermodynamics associated with the binding profile.

Other experimental considerations include “buffer mismatch”; this occurs when the buffer used to solubilise the ligand in the syringe is not the same as that used in the sample cell with the target receptor. This can result in artificial heat changes, therefore it is important to ensure the salt concentrations, pH and organic solvent concentrations are maintained between the two buffers of the syringe and cell samples.[210]

Organic solvents such as DMSO are often used to solubilise ligands for the purposes of biophysical analysis. As such, to minimise further chance of “buffer mismatch” it is important that the concentration of organic solvent within the syringe is identical to that of the target/ receptor within the sample cell.[210] In some cases there may be some incidents of interactions between the organic solvent and the target/ receptor, which may obscure the binding affinity determination. To take these organic solvent binding artefacts into consideration, the organic solvent can be added to the experimental standards discussed previously.

2.3.6 Binding affinity determination by ITC

General experimental procedures

The experimental procedure and the conditions used for the ITC experiments conducted in this section are discussed in detail in Chapter 6. However, briefly all unlabelled protein samples were prepared in the ITC buffer consisting of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ at pH 7.5.

The concentration of the protein and ligand were kept constant for the initial experiments of each inhibitor and subsequently varied as required to obtain an optimal binding curve. Due to the restricted solubility of the ligands the experimental process was reversed, meaning that the protein NCS1 was titrated into the cell contatining the desired Inhibitor.

2.3.6.1 Inhibitor1

The binding between NCS1 and Inhibitor 1 was investigated using ITC, numerous experiments were performed changing the ratio of protein and ligand, the resultant data has been summarised (Table 2.10).

Optimisation of the experiment was carried out, the resultant isotherm can be seen (Figure 2.45), analysis of the raw data from the experiment was conducted in the software package Origin7 (OriginLab, Northampton, MA). The data indicates an interaction between NCS1 and Inhibitor 1 as observed with the NMR data.

This experiment was carried out at the calculated concentrations of 400 μM NCS1 and 400 μM Inhibitor 1, however analysis of the raw data failed to produce a line of best fit of any statistical significance. It was noted during the experiment, that the solubility of the inhibitor in the aqueous buffer seemed limited (observed through precipitation of the ligand in aqueous solution), therefore it was postulated that the concentration of inhibitor in solution was significantly less than expected.

Therefore the raw data was re-analysed and fitted under this assumption and the concentration of Inhibitor 1 was adjusted accordingly, the best fitting of the curve occurred when the concentration of the Inhibitor was estimated as being 50 μM. The thermodynamic parameters for this interaction can be found in the appendix (Figure A.2.1) and are discussed in detail 2.3.6.3. To support the theory that the concentration of Inhibitor 1 in solution was indeed 50 μM, the corresponding experiment was conducted in duplicate, the thermodynamic parameters can be found in the appendix table A.2.1, the resultant isotherms found in Figure 2.46.

From the thermodynamic data of the duplicate experiments, which do not match that of the adjusted data, it can be concluded that the concentration of Inhibitor 1 in solution was not 50 μM and hence it's calculated K_D is invalid.

Table 2.10- ITC data of Inhibitor1. A table showing the thermodynamic properties from the ITC experiments of Inhibitor 1 with NCS1. The statistical errors of the ΔH parameter calculated with Origin Pro7 can be seen for each experiment and is indicative of a poor goodness of fit for each experiment. Therefore the calculated K_D is unreliable and inaccurate.

Experiment Title	ΔH (kcal mol^{-1})	TΔS (kcal mol-1)	ΔG (kcal mol-1)	K_D μM	Chi^2
NCS1 1 mM: Inhib 1 200 μM	-25.09 (±7876)	+18.63	-6.46	18.4	390300
NCS1 1 mM: Inhib 1 400 μM	-26.4 (±6328)	+20.00	-6.4	19.4	2760
NCS1 1 mM: Inhib 1 600 μM	-46.4 (±6328)	+40.53	-5.87	58.14	326200
NCS1 400 μM: Inhib 1 400 μM	-13.4 (±1116)	+6.2	-7.2	5.5	969500
NCS1 800 μM: Inhib 1 400 μM	-35.9 (±9948)	+30.10	-5.8	59.88	233300
NCS1 600 μM: Inhib 1 400 μM	-20.2 (±990.4)	+13.41	-6.79	15.26	65500

.

2.3.6.2 Inhibitor 2

The binding affinity of Inhibitor 2 with NCS1 was also investigated using ITC, as with Inhibitor 1 the solubility of the compound severely restricted the experimental technique. The initial experiment was carried out at a concentration of 1 mM NCS1: 100 μM Inhibitor 2, it is believed that even at the relatively low concentration precipitation of the inhibitor can be interpreted from the double heat spikes of the isotherm (Figure 2.47). The thermodynamic data can be observed in the appendix (Table A.2.2 and Figure A.2.2) a more detailed discussion can be found in 2.3.6.3. From this data it can be concluded that the solubility of Inhibitor 2 prevents determination of an accurate K_D.

2.3.6.3 Comparison of binding affinity determination between inhibitors 1 and 2 with NCS1

Broadly it can be observed that the thermodynamic properties of Inhibitor 1 and Inhibitor 2 share similarities in their profiles, both interactions appear enthalpically driven. As previously discussed in 2.3.5, ΔH is a measure of the number of intermolecular interactions formed or broken during a binding event and TΔS is related to the displacement of water from apolar surfaces and or change in conformational entropy through the binding interaction. In all experiments for Inhibitor 1 the ΔH upon binding of the inhibitor to NCS1 appeared to be moderate and negative and hence is the greator contributer to the binding affinity, whereas TΔS is smaller and positive, it can therefore be considered as insignificant.

With Inhibitor 2, although the ΔH upon binding is still large and negative and the TΔS is a positive value, it is much larger than those determined for Inhibitor 1 and so could not be considered as being insignificant.

The major limiting factor of both inhibitors with this experimental technique was their solubility, this meant that in many cases the goodness of fit chi-square values (Chi^2) were not minimal and hence determind binding affinity data was innacurate.[212–214] Therefore although it can be confirmed that there are interactions between the first generation of inhibitors and NCS1 as seen in the previously collected NMR data, determinaton of any affinity values using this technique are unreliable.

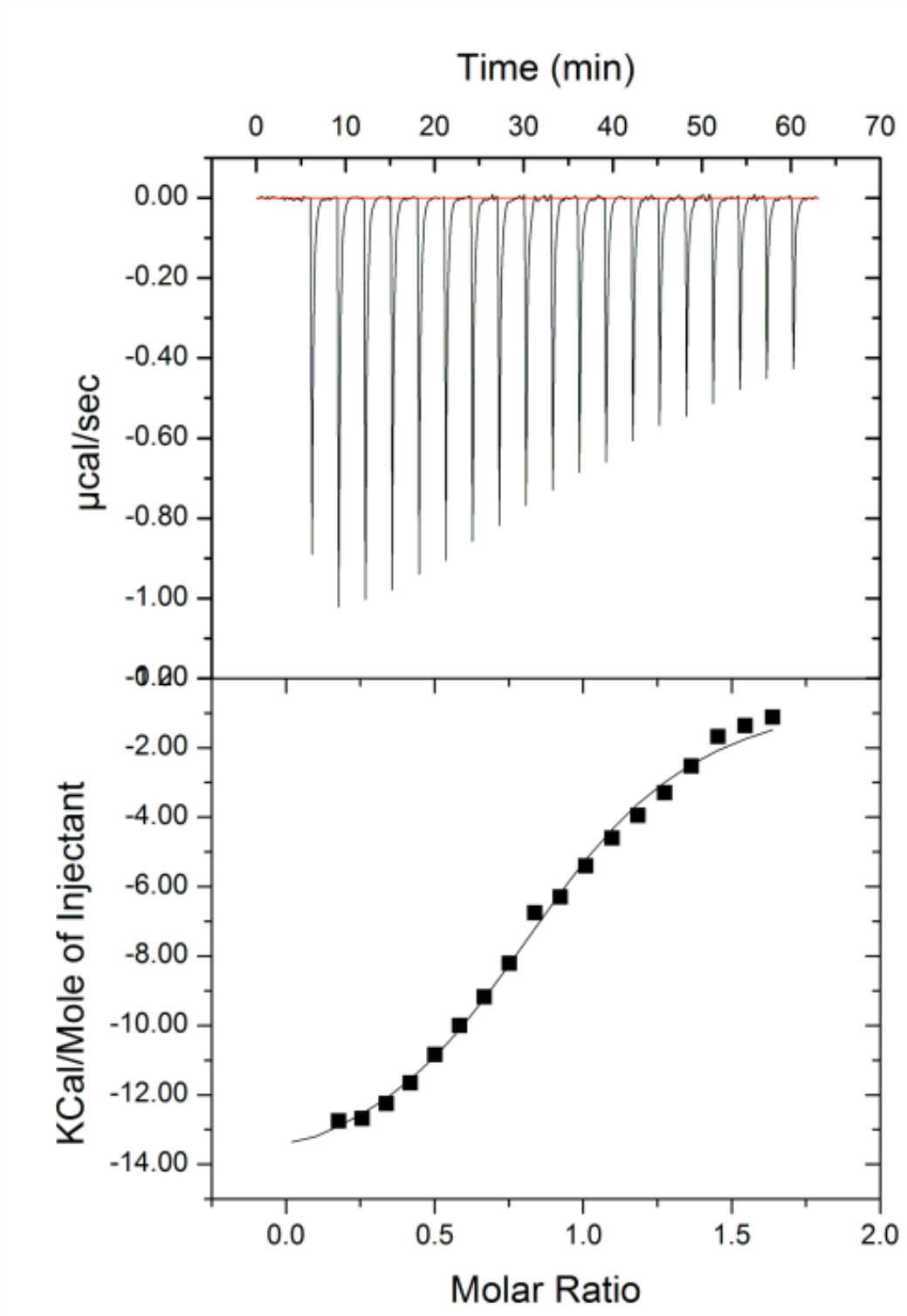


Figure 2.45- Binding affinity determination- Isothermal Titration Calorimetry Curve. ITC Isotherm top) and resultant curve (bottom) from the experiment of NCS1 400 µM titrated in to Inhibitor 1, 400µM at 25 °C, 1.34% DMSO, 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5. Fitting of the curve using Origin Pro7 assuming the concentration of Inhibitor 1 is 50 µM using the single set of sites curve fitting model produced a 1:1 binding model, a derived binding affinity K_D of 5.5 µM, ΔH= -13.4 Kcal mol^{-1}, $T\Delta S$ = 6.2 Kcal mol^{-1}, ΔG= -7.2 Kcal mol^{-1}, Chi^ 2= 969500.

.

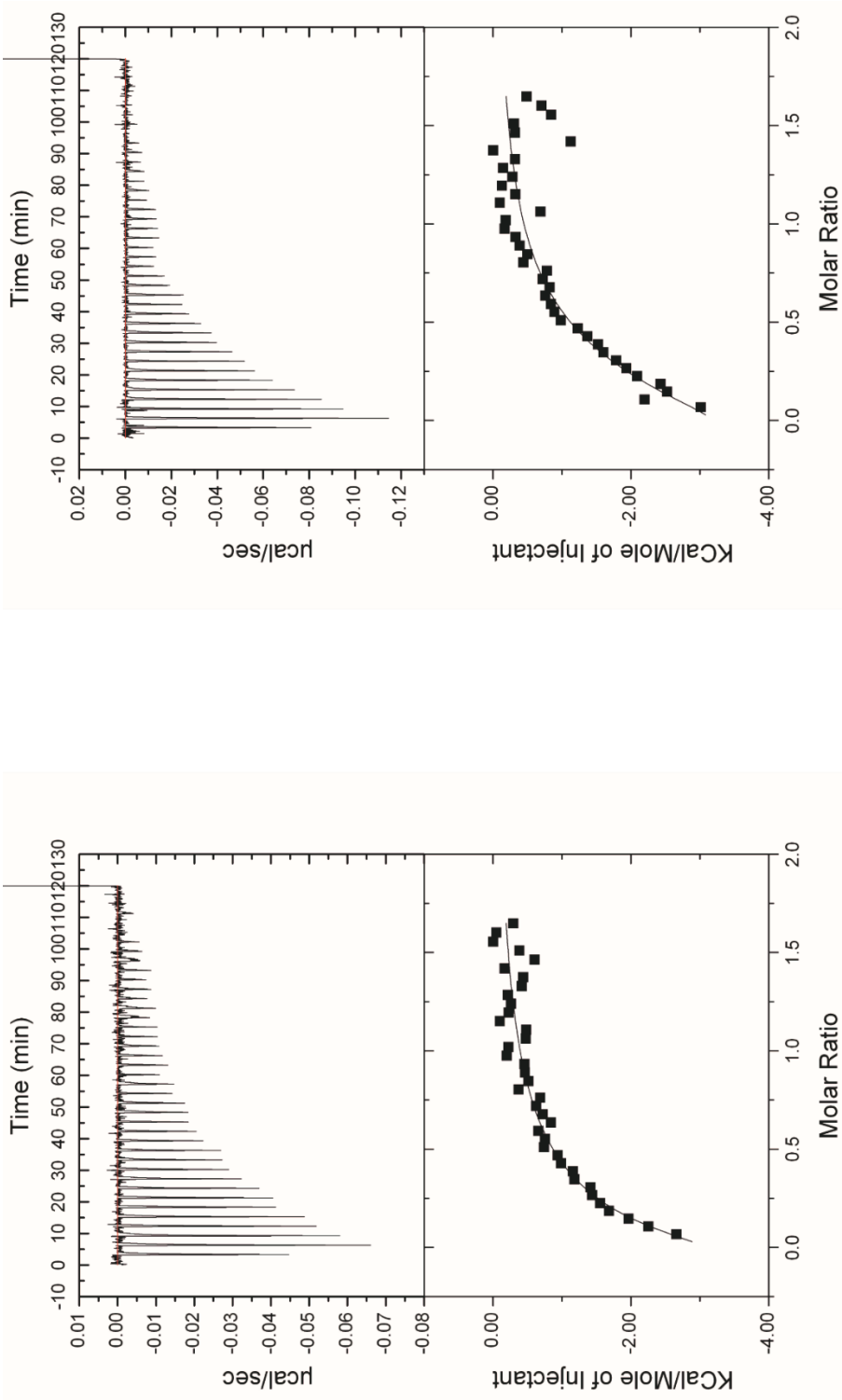


Figure 2.46- Binding affinity determination Isothermal Titration Calorimetry Curves. ITC Isotherm (top) and resultant curve (bottom) from the duplicate experiment of NCS1 400 µM titrated in to Inhibitor 1 50µM at 25 °C, 0.166% DMSO, 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5. Fitting of the curve using Origin Pro7 assuming the concentration of Inhibitor 1 is 50 µM using the single set of sites curve fitting model thermodynamic data can be found in the appendix Table A.3.1, however the Chi^2 values are 17230 and 68980 respectively.

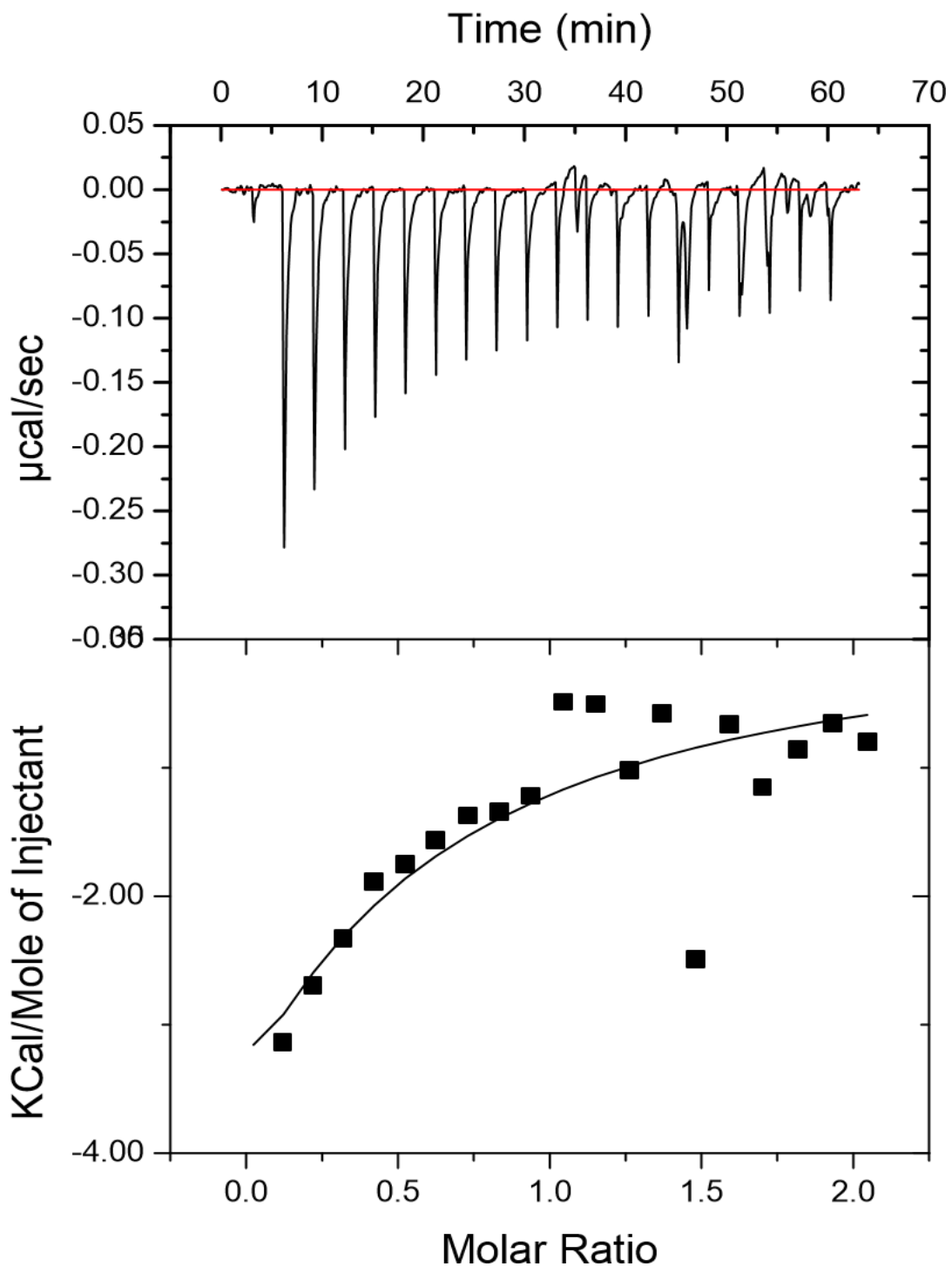


Figure 2.47- Binding affinity determination Isothermal Titration Calorimetry Curve for Inhibitor 2. ITC Isotherm (top) and resultant curve (bottom) from the experiment of NCS1 1 mM, titrated in to Inhibitor 2 100 µM at 25 °C, 0.34% DMSO, 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5. Double heat spikes seen in the top isotherm likely to have occurred due to the limited solubility of Inhibitor 2 resulting in precipitation. Fitting of the curve was conducted in Origin Pro7 using the single set of sites curve fitting model produced a binding model with a derived binding affinity K_D of 179.5 µM undesiarable enthalpic values as follows ΔH= -4.708 E8 Kcal mol^{-1}, -TΔS= 4.71 E8 Kcal mol^{-1}, ΔG= -9.42 E+05 Kcal, mol^{-1} Chi^2= 252000.

2.3.7 Introduction to Fluorescence Spectrophotometry

The emission of light from any given surface is known as luminescence and can be split into two categories, fluorescence and phosphorescence, depending on the excited state.[216] Here though we shall discuss fluorescence, in particular the use of intrinsic fluorescence for investigation of the binding interactions of a protein. Intrinsic protein fluorescence is a product of the aromatic sidechain residues of the amino acids tryptophan, tyrosine and phenylalanine within the protein itself, rather than a fluorescence label.[216]

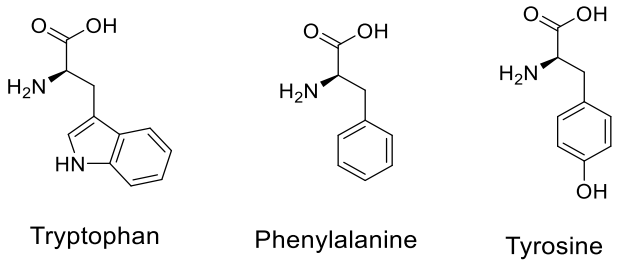


Figure 2.48- Chemical structures of the three aromatic sidechains of the amino acid residues tryptophan, phenylalanine and tyrosine. Of which tryptophan is the most powerful fluorescence tool as it is stronger in emission than tyrosine or phenylalanine and is extremely sensitive to the local environment that surrounds the indole sidechain.

In a fluorescence experiment, the sample is excited at a specific wavelength of light relating to aromatic residue of interest; for example, in tryptophan fluorescence the excitation wavelength is 285 nm and the emission wavelength for detection is between 300 and 350 nm. Tryptophan fluorescence is highly sensitive to the local environment that surrounds the tryptophan's indole sidechain; hence tryptophan fluorescence can be used to detect conformational changes of a protein at low concentration.[216,222]

These changes in conformation present themselves in two ways. A blue shift in the fluorescence (a shift to the left of the native fluorescence i.e. less than 300 nm) is observed in those circumstances where the tryptophan residues become more buried into the hydrophobic surface upon interaction of the protein with a particular ligand or binding partner.[216] On the other hand, a red shift in the tryptophan fluorescence (a shift to the right of the emission spectra i.e. above 350 nm) occurs when the tryptophan residues become more exposed and this is a possible indication of protein unfolding.[216]

The tryptophan fluorescence data collected for this thesis was carried out on a Cary Eclipse fluorescence spectrophotometer (Agilent technologies), the raw data was analysed in Origin7 (OriginLab, Northampton, MA).

2.3.8 Binding affinity determination using Fluorescence Spectroscopy

General experimental procedures

All fluorescence data of the first generation of inhibitors was collected by Dr Liam Dorr at The University of Liverpool on a Cary Eclipse Fluorescence spectrophotometer (Agilent technologies). The raw data was analysed in Origin7 (OriginLab, Northampton, MA) by myself.

Unlabelled NCS1 was prepared as described in Chapter 6. Briefly, unlabelled NCS1 was buffer exchanged into fluorescence buffer consisting of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5; and aliquoted into 17 samples at a final concentration of 1 μM (v/v 200 μL).

For the individual single point reads of the tryptophan fluorescence assay, each inhibitor was added in the required volume from a single stock solution in 100% DMSO, to the already aliquoted NCS1 samples contained in eppendorf microfuge tubes. The concentration range for each inhibitor over the 17 samples was 0 - 90 μM. This process was repeated for all inhibitors. Each individual assay sample was transferred to a quartz 16.160-F/Q/10 cuvette (Starna Scientific) and excited at a wavelength (λ) of 280 nm, with the excitation slit width 5 nm and the emission slit width 20 nm.

2.3.8.3 Inhibitor binding assay

Intrinsic fluorescence is highly sensitive to a change the local environment of tryptophan residues, the emission spectrum of tryptophan changes in response to conformational changes in the local environment.[271] This makes tryptophan fluorescence a useful tool for the investigation of changes in a proteins structure, whether it be due to denaturation, interactions with a ligand, or a known binding partner, it can also be used to determine binding affinities.[272]

Tryptophan fluorescence was used to investigate the interactions of NCS1 with inhibitors 1, 2 and 3 as NCS1 is known to contain two tryptophan residues in the hydrophobic binding cleft Trp30 and Trp103. Experimentally the required concentration of the protein and the ligand is extremely small in comparison to that of NMR and ITC techniques. The changes in tryptophan fluorescence of NCS1 upon the addition of 0 – 90 μM inhibitors 1, 2 and 3 can be seen (Figure 2.49). Immediately from the plots of the raw data it is clear that each inhibitor has an effect on the tryptophan fluorescence of NCS1.

In the first instance, increasing the concentrations of Inhibitor 1 (Figure 2.49 a) caused a quenching of the tryptophan emission.

The effects of increasing the concentrations of Inhibitor 2 upon NCS1, appear to cause a blue shift in the tryptophan fluorescence after 7.5 μM. This change in the fluorescence emission, indicates that the indole ring of the tryptophan residue has become more buried within the cleft.[216]

At increasing concentrations of Inhibitor 3, NCS1 appears to undergo the opposite change in structure as was caused by Inhibitor 2. At 10 μM the intrinsic tryptophan fluorescence emission of NCS1 undergoes a red shift, this indicates that the tryptophan is now more.[216]

The tryptophan fluorescence assays were used as an additional method to investigate the effects of the inhibitors on NCS1; however, no further determination of binding affinity was carried out using this technique due to the additional effects of red/blue shifts which complicated the analyses. Instead, the fact that there are effects on the tryptophan fluorescence by all 3 inhibitors suggest that that each on is interacting with NCS1 albeit differently.

.

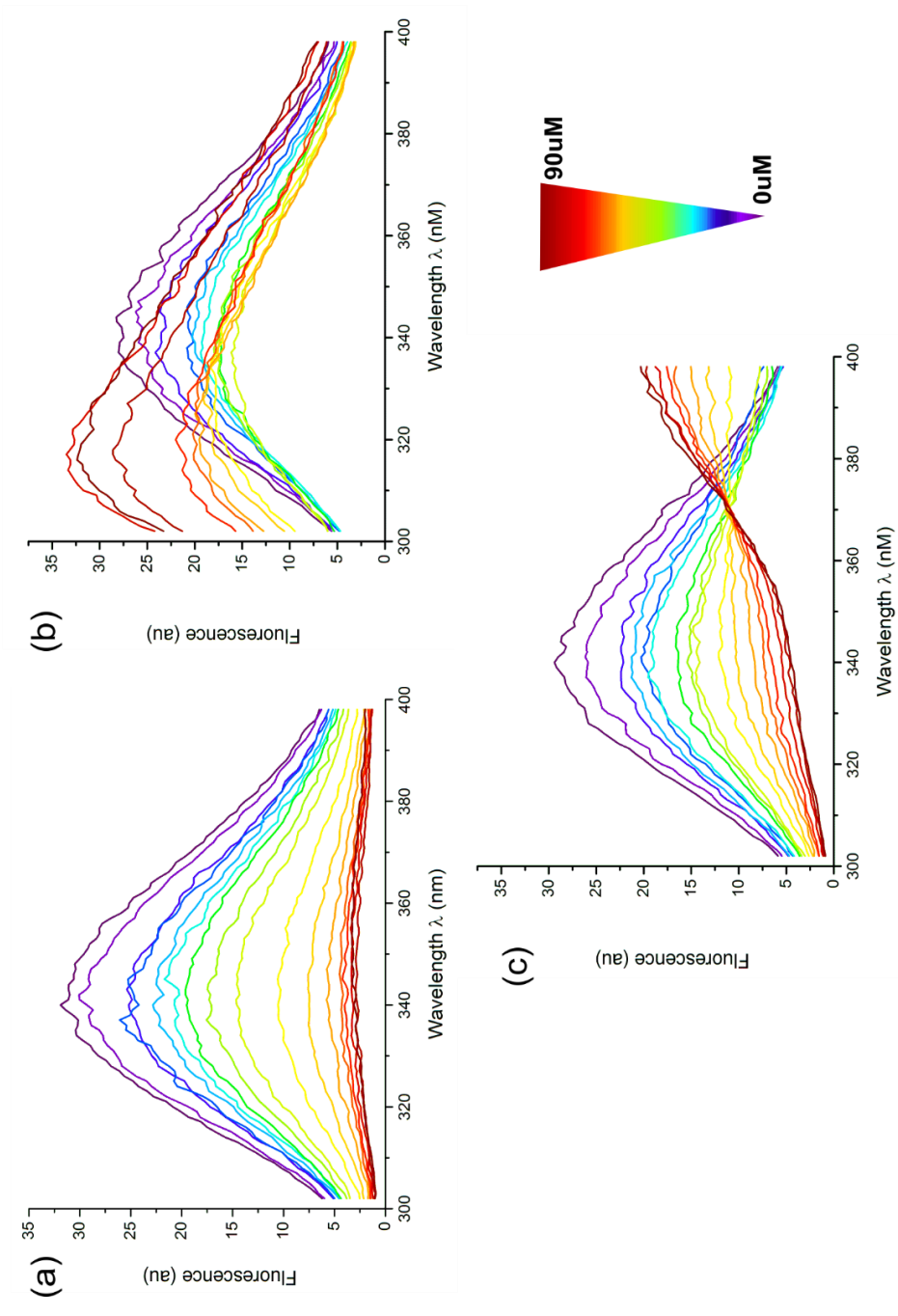


Figure 2.49- Fluorescence binding titration assays of NCS1 1 µM with inhibitors 1, 2 and 3 over a range of 0 – 90 µM. (a) Inhibitor 1 fluorescence binding assay, NCS1 at 1 µM and Inhibitor 1 titrated in to the individual protein samples at concentrations 0 µM, 0.5 µM, 1 µM, 1.5 µM, 2.0 µM, 2.5 µM, 5 µM, 7.5 µM, 10 µM, 20 µM, 30 µM, 40 µM, 50 µM, 60 µM, 70 µM, 80 µM, 90 µM. **(b)** Inhibitor 2 fluorescence binding assay, NCS1 at 1 µM and Inhibitor 2 titrated in to the individual protein samples at concentrations as described for a. **(c)** Inhibitor 3 fluorescence binding assay, NCS1 at 1 µM and Inhibitor 3 titrated in to the individual protein samples at concentrations as described for a and b.

2.4 Summary

The first generation of inhibitors designed to modulate the interaction between NCS1 and D2R peptide, were analysed extensively using numerous biophysical techniques including; NMR spectroscopy, ITC and tryptophan fluorescence spectrophotometry.

All three methods of analysis indicated an interaction between inhibitors 1, 2 and 3 with NCS1, however limitations brought about due to their limited solubility in aqueous buffer at concentrations needed for these techniques meant that affinity of binding was unable to be determined accurately. Therefore it was decided to develop a second generation of inhibitors that would exhibit an improved solubility profile.

The approach towards the second generation compound is shown in the computational section (3.1) and the synthetic approach (Section 3.2). The second generation of inhibitors were designed based on the structure of Inhibitor 2, from the NMR data Inhibitor 1 appeared to cause too much of a change in the secondary structure of NCS1 at lower concentrations.

The effects of Inhibitor 3 upon the NCS1 residues at the same concentration as Inhibitor 1 and Inhibitor 2 (500 μM) did not appear to have an obvious effect on the targeted NCS1 residues, this along with the fluorescence data, that possibly indicated unfolding events at higher concentrations of Inhibitor 3, made it an unsuitable candidate for further development.

Therefore Inhibitor 2 was selected, its NMR data indicated the possibility of a binding interaction with NCS1 and this was better observed than with Inhibitor 1 whose effect of severe peak broadening and attenuation meant that analysis of the residues perturbed by any possible interaction was not accurate. The fluorescence data obtained from Inhibitor 2, also indicated a change in the accessibility of the tryptophan residues within the hydrophobic binding pocket, the lack of accurate ITC data obtained due to its limited solubility made Inhibitor 2 an obvious candidate for further structural development.

Upon the synthesis of the second generation of inhibitors, the biophysical techniques to be applied for investigation of the interactions with NCS1 would be the same techniques as used for the first generation. A detailed discussion of the results of this analysis can be found in Chapter 3 Structure-based drug design: The Second Generation Section 3.3 and includes an investigation using NMR techniques into the solubility of both the first generation of inhibitors compared to the second generation.

Chapter 3

Structure-based drug design: The Second Generation

3 Structure-based drug design: The Second Generation

3.1. Second Generation: Computational design

3.1.1 Introduction

Following the design and synthesis of the first generation of inhibitors (Chapter 2, sections 2.1 and 2.2) and the results of the biophysical analysis (Section 2.3), it was apparent that a major limiting factor in the first generation analogues was their very limited solubility. Therefore the most promising inhibitor from the first generation was selected following a detailed analysis of the predicted binding poses from the computational modelling, rather than from experimental binding affinities. From the resulting scaffold, a number of analogues were designed with solubilising groups incorporated into the scaffold. Inhibitor 2 was selected due to the consistently predicted binding pose along the hydrophobic groove of NCS1. Its selection was also in part due to the non-specific binding effects of Inhibitor 1 on NCS1 as observed from the NMR chemical shift perturbation experiments. There was also a lack of observable binding with Inhibitor 3.

Visual analysis of the single top binding pose predicted for Inhibitor 2 (Figure 3.1), indicated the location of a number of moieties within Inhibitor 2 in proximity to the side chains of NCS1. From the computational docking we observed the methyl group was found to be orientated towards the phenyl ring of Phe48 at a distance of 4.2 Å (Figure 3.1) and the position of the *para* methyl group on the pyridine ring was 1.8 Å from the phenyl ring of Tyr52. The ortho methyl group of the pyridine ring was situated 3.1 Å from Ser184, with the methoxy group appearing to be 2.9 Å from Trp103. Finally the methylene CH_2 indicated by the blue box (Figure 3.1) was believed to orientate the aromatic ring into a position at the top of the binding groove around Leu43 at a distance of 3.5 Å.

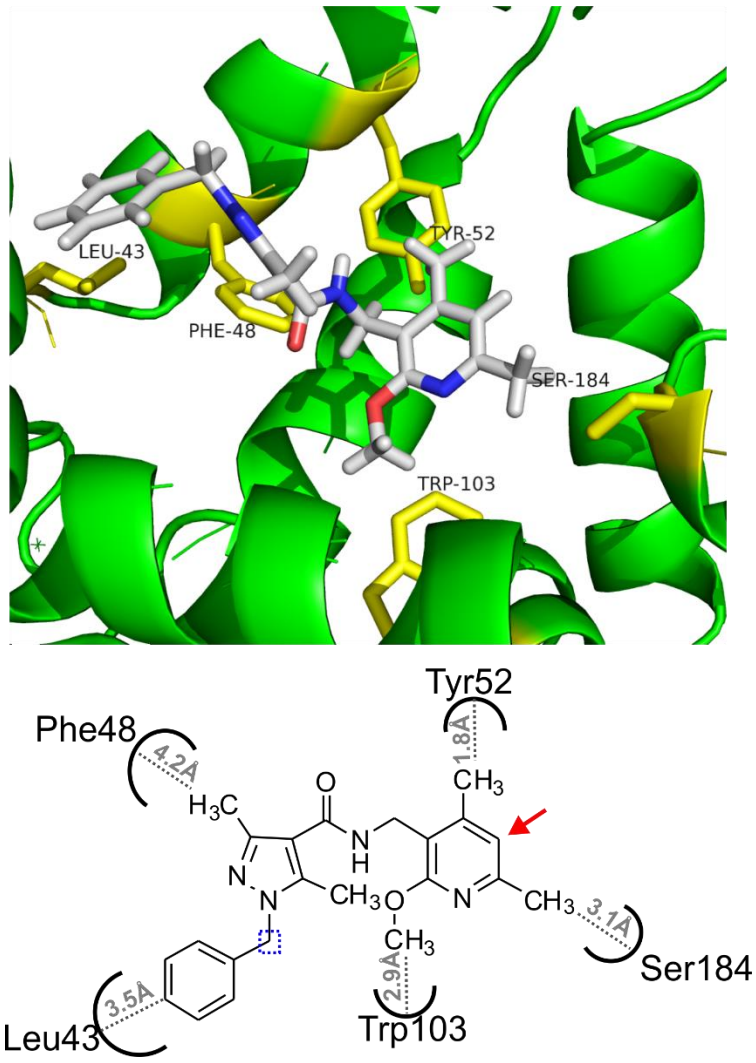


Figure 3.1– Optimisation of Inhibitor 2. Visual analysis of the top binding pose of Inhibitor 2 in the hydrophobic binding cleft of NCS1 was used to analyse any positions on the scaffold that could be optimised by the addition of a solubilising group. The distances between the residues of NCS1 and the functional groups of Inhibitor2 are indicated in grey. The methyl group was found to be orientated towards the phenyl ring of Phe48, at a distance of 4.2 Å. Position of the methyl group found to be 1.8 Å from the phenyl ring of Tyr52. Methyl group situated 3.1 Å from Ser184. The methoxy group appears to be 2.9 Å from Trp130. The methylene CH_2 (blue box) orientates the aromatic ring, which is situated at the top of the binding groove around 3.5 Å from Leu43. The red arrow represents an accessible position where a solubilising group could be incorporated into the scaffold, which is solvent exposed free space.

3.1.2 Adapting the modelling pipeline: Computational optimisation techniques

Based on the model described above it was apparent; that the unsubstituted position on the pyridine ring in inhibitor 2, would be the most suitable position for the addition of a solubilising group as it would be highly solvent exposed (red arrow Figure 3.1). The phenyl ring situated at the top of the binding groove, appears to anchor the ligand in position through interaction with Leu43. Therefore following the Topliss scheme for optimisation of phenyl ring substitution, the only substitutions suggested in this region of the scaffold was the addition of a 4-chloro/ para-chloro group.[273] Despite

the more lipophilic and electron withdrawing nature of the 4-chloro substitution, the ease of synthesis relative to other substituted compounds makes it a good first choice for optimisation processes.[273] Furthermore the *ortho* and *para* substituted methyl groups on the pyridine ring were removed, due to the fact that their presence could render the molecule more lipophilic and thus less soluble.[274] The *ortho* methoxy group of the pyridine ring was removed in favour of either a pyridine or pyridone ring which would allow for substitutions at both *meta* positions. Finally selection of a suitable solubilising group, resulted in investigations in heteroaliphatic and heteroaromatic rings, to determine which ring type would be most suitable.[275] The different types of rings chosen to be screened (Figure 3.2) included; *N*-methyl piperazine, morpholine, piperadine, imidazole and 1, 2, 4 triazole.

With the structural analysis of Inhibitor 2 complete, four main scaffolds were designed (Figure 3.2), (a) and (b) pyridine, or (c) and (d) pyridone analogues, with *meta* substitutions (R) being either amide or amine. This resulted in the design of 72 analogues however, due to the lack of commercial availability of a number of the solubilising groups, only 54 of the 72 analogues of Inhibitor 2 (Appendix Figure A.3.1 - A.3.3) were subsequently built and energy minimised in Spartan.[225]

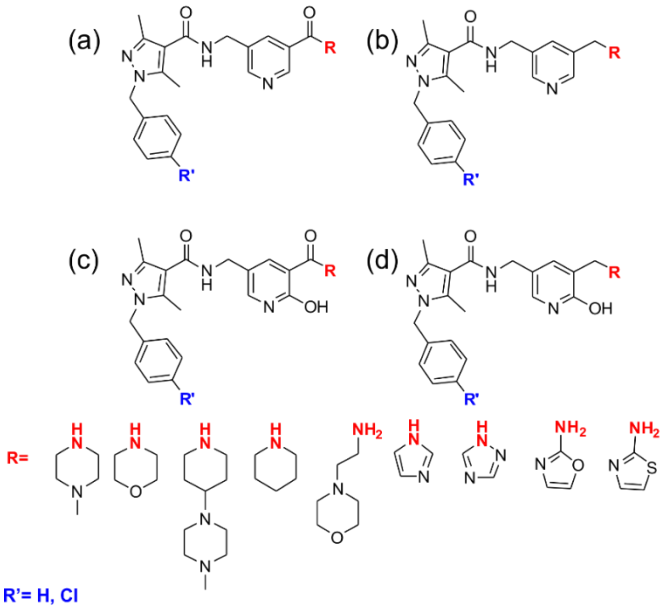


Figure 3.2- Inhibitor 2 analogue design. The four main scaffolds can be observed with substitutions at the R (red) and R' (blue) positions, resulted in 54 possible analogues of Inhibitor 2 to be computationally screened.

As with the workflow of the first generation of compounds, the 54 analogues of Inhibitor 2 were docked into NCS1 using GOLD and their predicted binding interactions assessed scored using Goldscore.[99] The one hundred docking poses of each of the compounds, were subsequently re-scored using the three scoring

functions, Astex statistical potential (ASP), Chemscore and Chem Piecewise Linear Potential (PLP).[99,111] An average of the scoring functions was taken for each compound and the ADMET profiles (the physiochemical properties) were calculated in Pipeline Pilot.[125] Despite the additional solubilising groups the solubility profile of the 54 compounds was not always ideal (Appendix Section A.3 Table A.3.1). The predicted AlogP ranged from the unfavourable 3.787 to a much more favourable value of 0.413, which is considerably improved when compared to that of Inhibitor 2 whose AlogP was calculated as 3.737. The molecular solubility was also calculated for the 54 molecules and values ranged from the favourable -4.42 to a much less favourable value of -7.95, considerably less than that of Inhibitor 2 whose molecular solubility was calculated as -6.47. In fact of the 54 compounds designed to have an improved solubility, 15 were found to have a poorer profile than the parent compound and a further 19 were more negative than -6.0.

Due to the solubility issues of the first generation of inhibitors and the subsequent challenges that this caused with the *in vitro* binding studies, it was decided the selection process should filter out compounds with poor calculated molecular solubility. Compounds with a molecular solubility more negative than that of Inhibitor 2 were removed, the final 39 were ranked in ascending order of molecular solubility and the top 12 compounds (Table 3.1 and Figure 3.4) were selected for retro-synthetic analysis to determine compounds to synthesise.

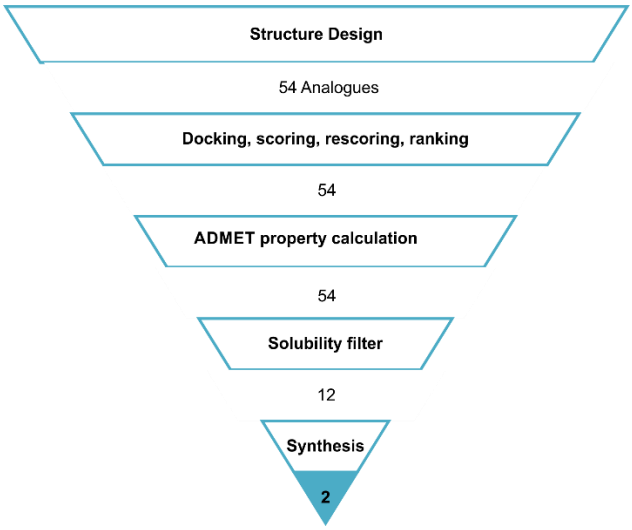


Figure 3.3 - Second Generation Computational workflow. The computational workflow applied to the Inhibitor 2 analogues adapted from the previous first generation computational workflow used to select inhibitors 1, 2 and 3 (Chapter 2, Section 2.1). 54 compounds were designed and built in the molecular modelling package Spartan, they were subsequently docked into NCS1 using the docking programme GOLD where their binding poses were scored and re-scored. The ADMET profiles of each of the 54 compounds were then calculated and the ligands were ranked in order of their solubility and the top 10 compounds were selected for retro-synthetic analysis.

Table 3.1- Table showing the ADMET properties for the top 12 molecule derivatives on Inhibitor 2. Compounds are listed in descending order with respects to the overall best solubility profile. The molecular weight and solubility functions are indicated using a traffic light system following the Lipinski parameters and the filter of molecular solubility of Inhibitor 2 as -6.0: good values (green), intermediate (amber) and poor values (red).

Compound 3.13 has been included in the table in italics as it was selected for synthesis despite not ranking amongst the top 12 compounds.

D2R peptide Ligand Efficiency = 0.061

Name	Molecular Solubility	Number H Acceptors	Number H Donors	ALogP	Molecular Weight	Average Goldscore	Ligand Efficiency
3.1	-4.42	5	2	1.08	448.56058	39.61	0.088
3.2	-4.692	5	2	0.811	435.51876	37.94	0.087
3.3	-5.052	5	1	1.503	446.5447	39.18	0.087
3.4	-5.202	5	2	1.745	483.00564	41.44	0.0857
3.5	-5.292	5	1	1.234	433.50288	38.12	0.087
3.6	-5.372	5	1	2.163	432.56118	39.20	0.090
3.7	-5.397	6	1	1.614	529.67635	47.58	0.089
3.8	-5.472	5	2	1.475	469.96382	47.04	0.100
3.9	-5.585	6	2	1.855	566.13729	39.70	0.070
3.10	-5.64	5	1	1.893	419.51936	36.70	0.087
3.11	-5.695	6	2	1.166	476.57068	47.08	0.098
3.12	-5.732	4	2	2.04	433.54594	39.43	0.090
3.13	*-6.066*	*5*	*1*	*1.898*	*467.950*	*37.36*	*0.079*

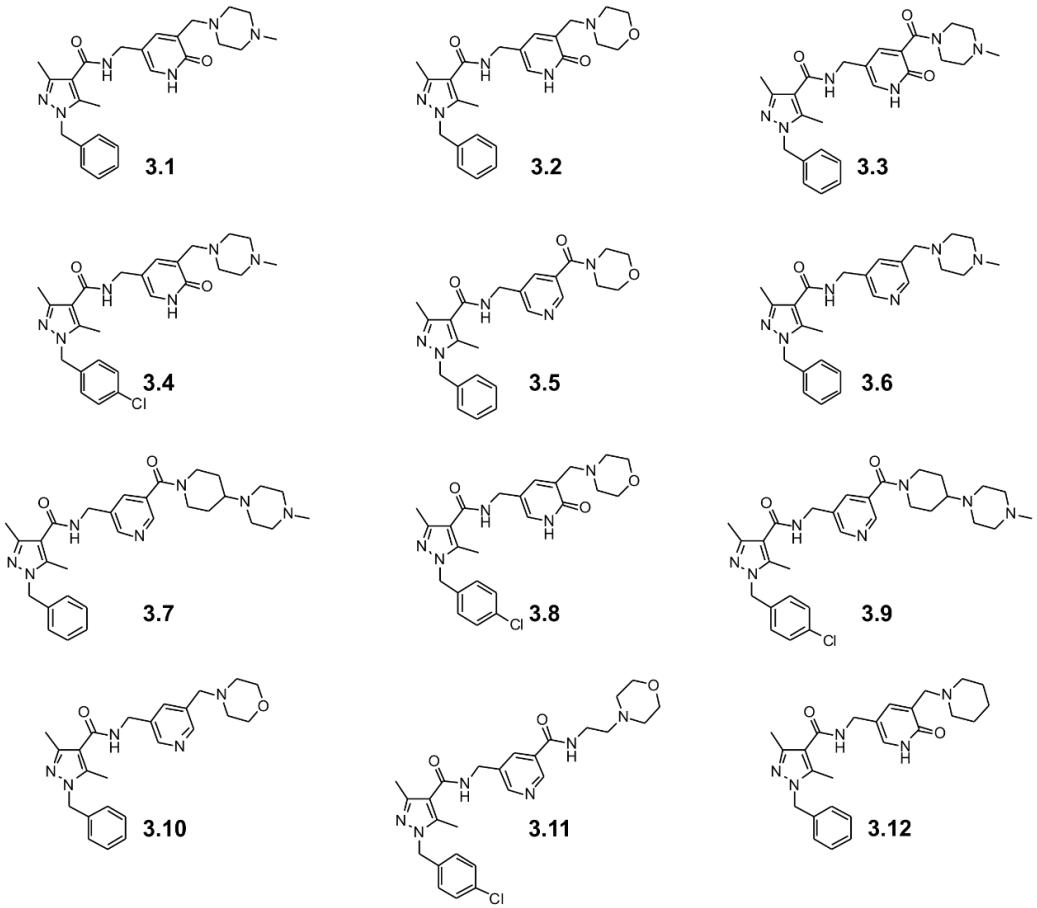


Figure 3.4- Top 12 compound derivatives of Inhibitor 2. The compounds ranked in order of their solubility profile. The relative ADMET properties for each compound can be seen in table 3.1.

After retrosynthetic analysis, **3.5** was selected for synthesis and the synthetic route outlined in Section 3.2. Due to challenges encountered during the synthesis it was decided to also synthesise **3.13**, a compound closely related to **3.5**, the two only differ by a para-substituted chlorine (Figure 3.5). Compound **3.13** was not within the top 12 compounds with the best molecular solubility, however its predicted solubility was still predicted as being better that Inhibitor 2 (Table 3.1). The top predicted binding pose for both **3.5** and **3.13** (Figure 3.5) indicate that the top binding pose of **3.5** is different to that of Inhibitor 2, in comparison to **3.13** which is predicted to bind in a similar manner down the groove albeit with the opposite orientation.

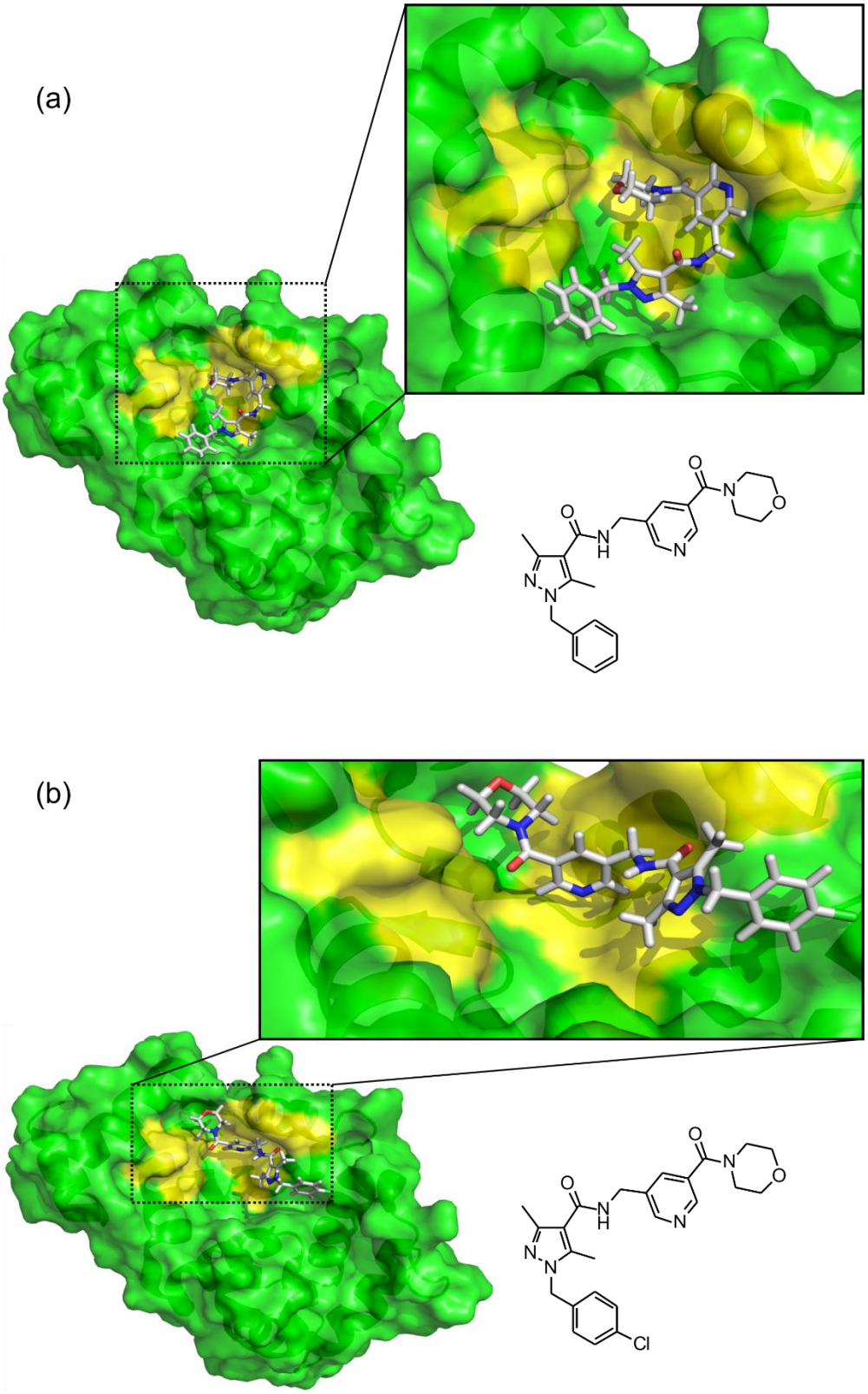


Figure 3.5- Predicted binding poses of 3.5 and 3.13. (a) The top predicted binding pose of **3.5** with NCS1 (green) and the key residues highlighted (yellow) [4] along with the chemical structure. Compound **3.5** appears to adopt a curved binding pose in the lower region of the key residues within the hydrophobic binding groove of NCS1. **(b)** The top predicted binding pose of **3.13** and NCS1 (green) with the key residues highlighted (yellow) along with the chemical structure. Compound **3.13** appears to adopt a similar binding pose to that of Inhibitor 2 along the hydrophobic binding cleft, however the orientation is reversed.

3.1.2 Summary

The significantly reduced solubility profiles of each of the three inhibitors determined during the biophysical analysis (Chapter 2.3) led to the subsequent development of a second generation of inhibitors.

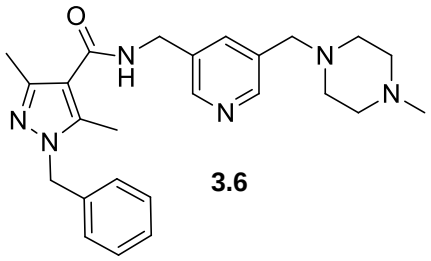
Developing the second generation of inhibitors was conducted using Inhibitor 2 as a template due to its biophysical binding results (Chapter 2.3). The computational approach involved a more manual design process informed by docking results, through the analysis of the predicted binding interactions of Inhibitor 2. Compound design was informed by medicinal chemistry knowledge of solubilising groups. The 54 generated compounds were docked and their physiochemical properties calculated as with the first generation of inhibitors; however, the solubility profile of each compound was a key filter in the second generation selection process. The solubility profiles of the compounds were compared to that of Inhibitor 2 and those whose profile was not significantly improved were discarded. The remaining compounds were listed in ascending order with respect to their molecular solubility's and the top 12 compounds were selected for retrosynthetic analysis. Compound **3.5** was intended to be the initial synthetic target, with the aim of synthesising a number of closely related analogues; however, due to the challenging synthetic route, **3.13** was the only compound suitable. This compound has a similar structure and predicted binding pose to that of Inhibitor 2 and, hence, deemed a suitable second generation compound.

It has been said that the road to a successfully developed therapeutic is long, expensive and often frustrating.[117] There are many different options when it comes to "how to best design" a compound or target a specific interaction and they are all valid in one way or another. However the pipeline that has been developed here is a rationale and efficient method that combines different computational techniques, to produce the desired outcome of identifying hit compounds. The methodology could be applied to other protein-protein interactions and could significantly reduce the cost associated with experimentally screening vast chemical libraries. Although these are only calculated predictions and the experimental results may be found to differ, it is rational to start with a targeted approach.

3.2 Second Generation Inhibitor Synthesis

3.2.1 Introduction

The first generation of inhibitors designed to disrupt the interaction between NCS1 and D2R peptide were analysed extensively using numerous biophysical techniques including NMR spectroscopy, ITC and tryptophan fluorescence (Chapter 2.3). However, limitations due to their limited solubility in aqueous buffer meant there were no quantifiable biophysical results. Therefore, it was decided to develop a second generation of inhibitors that would exhibit an improved solubility profile without losing efficacy towards the target. The selection process involved analysis of the predicted binding poses from the computational analysis of the first generation compounds; this was necessary since despite the extensive NMR experiments, there was a lack of accurate data that identified the exact binding locations of the inhibitors within NCS1. Inhibitor 2 was selected as the target scaffold because of the predicted binding pose along the hydrophobic groove of NCS1 (Chapter 2.1.5 Figure 2.13 d). As discussed previously, the computational process in designing and selecting these new generation analogues involved the incorporation of solubilising groups such as *N*-methylpiperazine, morpholine, piperidine, imidazole and 1,2,4-triazole. The pipeline took the 54 designed compounds and filtered them based on their solubility profile (Section 3.1). This resulted in 12 hit compounds that were analysed retrosynthetically to find a common route to synthesis that could be applied to all candidates. This would allow synthesis of a number of ligands with identical core scaffolds, but different solubilising functional groups. The first compound of the second generation, selected for synthesis and for the development of the synthetic route was **3.6**, due to the commercial availability of the reagents.

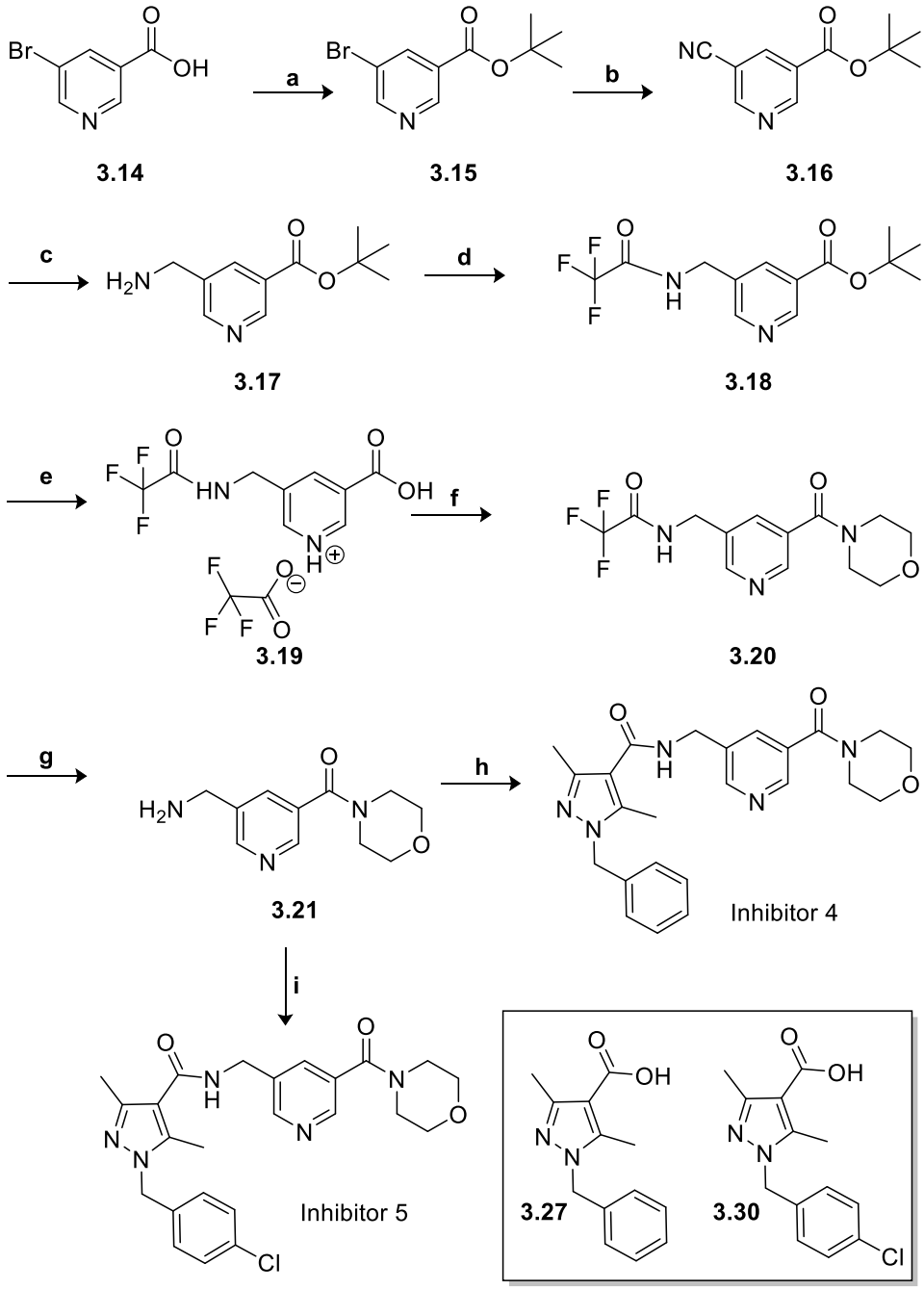


3.6

Figure 3.6- Initial synthetic target **3.6.**

3.2.2 Synthesis of Inhibitors 4 and 5

The overal synthetic routes used in the synthesis of of the two second generation compounds Inhibitors 4 and 5 can be seen in Scheme 3.1 and 3.5. The first five steps in the synthetic approach a-e **3.14 – 3.19** were developed to be applicable to seven of the original 12 molecular targets.

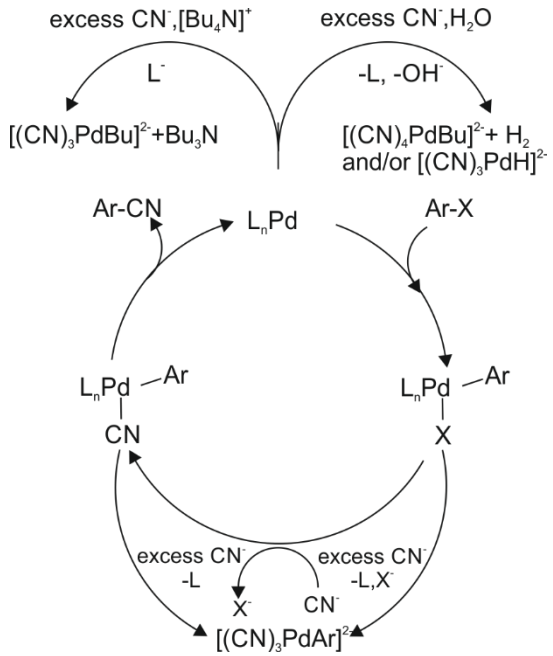


Scheme 3.1- Synthetic route to Inhibitors 4 and 5. (a) Tertiary butanol, EDCI, DMAP, 25 ˚C 72 hrs. 57%. **(b)** $Pd_2[dba]_3$, 1,1'-Bis(diphenylphosphino)ferrocene, anhydrous dimethylacetamide, $ZnCN_2$, Zn, 90 ˚C, 1 hr. 85%. **(c)** 25% NH_3, raney nickel, EtOH:THF 1:1, H_2, 25 ˚C 3 hrs. **(d)** Trifluoroacetic anhydride, pyridine, DCM, 25 ˚C 16 hrs. 68%. **(e)** Trifluoroacetic acid, DCM, 25 ˚C, 3 hrs. **(f)** HATU, anhydrous DMF, DIPEA, Morpholine, 25 ˚C, 96 hrs. 55%. **(g)** 7M methanolic ammonia, 25 ˚C, 96 hrs, 31%. **(h)** HATU, anhydrous DMF,DIPEA, **3.25**, 25 ˚C, 96 hrs. 36%. **(i)** HATU, anhydrous DMF, DIPEA, **3.28**, 25 ˚C , 96 hrs. 34%.

The initial step in the synthesis was *t*-butyl protection of the carboxylic acid functionality **3.14**; this was followed by cyanation of the aryl bromide **3.15** at the 5 position. Cyanation of aryl halides has previously been conducted using microwave irradiation and transition metal chemistry, with the displacement of halides from aromatic compounds using cheap and readily availiable cyano sources such as potassium cyanide or sodium cyanide.[276][277] Palladium (Pd) catalysts were selected due to their ability to tolerate a variety of different functionalities and conditions.[278] Microwave reactions require quite harsh conditions including high temperatures and pressures and the isolation of the product can be difficult resulting in low yields.[276]

Initially cyanation reactions were conducted prior to *t*-butyl protection using copper cyanide in acetonitrile and the microwave reactions heated to 160 ˚C; however no products were isolated. This was thought to be due to the difficulties with isolation of the pyridine carboxyllic acid, so protection of the carboxylic acid was then conducted as an initial step. However, it was thought that the high temperatures and pressures of the experimental conditions of the microwave reaction, caused the *t*-butyl group to be cleaved. This once again revealed the carboxylic acid and no product could be isolated.

Therefore, it was decided to investigate cyanation of *t*-butyl protected nicotinic acid using palladium catalysts. The *t*-butyl ester was introduced due to its stability to basic conditions and its lability under strongly acidic conditions, the reaction was successful yielding the product compound **3.16** with 57% yield.[279] The mechanism of the cyanation reaction using palladium catalysts was originally postulated by Takagi *et al.* in 1976. Mechanistic studies indicated that two cycles were involved in the reaction whereby one fed the other.[280] The first cycle was thought to occur like that of a typical Pd cross coupling reaction; in the second the Pd was the cyanide carrier.[280] More recently in an extensive review article, Anbarasan *et al.* suggested a similar catalytic cycle (Scheme 3.2). Oxidative addition into the aryl halide by the palladium species forms a Pd(II) aryl halide complex. This would then be followed by anion exchange replacing the halide and forming the Pd (II) cyano complex.[281] The final step in the catalytic cycle would be reductive elimination, forming the corresponding aryl cyano compound and re forming the active Pd(0) compound.[281]



Scheme 3.2– The proposed catalytic cycle of cyanation including deactivation pathways suggested by Anbarasan *et al.* adapted from [281].

Sources of cyanide and suitable solvents for the reaction were thought to be the limiting factors and were dependent upon the respective solubility. It was found that the higher the solubility of the cyanide salt the lower the reaction yield.[281,282] According to the original proposed mechanism, it was thought that problems with the reaction were due to the presence of water or excess cyanide, subsequently resulting in catalyst de-activation, both of which can be seen as side reactions in Scheme 3.2.[280,281] Water is believed to generate HCN which can lead to the formation of the inactive Pd (II) cyano complex. It has also been postulated that it improves the solubility of the sources of cyanide in organic solvent.[283,284] Excess cyanide was thought to inhibit the catalytic cycle through the interaction of the cyanide ions with the Pd (II) species, forming an inactive Pd (II) cyano complex that could not be reduced back to the active Pd (0) species.[280] This could be due to the strongly binding cyanide anions and so resulting in the termination of the reaction.[280,281] However, this theory was found to not necessarily be the case in many of the reaction screens that were tested in this synthesis.

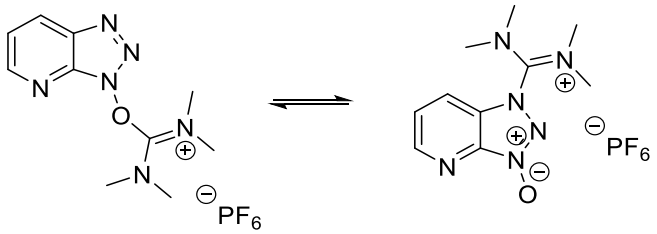
Two types of palladium ligands were used, tris (tribenzylideneacetone)di-palladium(0) $Pd_2(dba)_3$ and bis(dibenzylideneacetone) palladium(0) $Pd(dba)_2$ along with the ligand 1,1'-Bis(diphenylphophino)ferrocene (DPPF) to complete the catalytic system. The cyanide source used was zinc cyanide in conjunction with zinc powder, which was used to keep the palladium in the active oxidation state of 0. Equimolar equivalent were used in the majority of the reactions and in all cases the zinc cyanide used was in excess with respect to the palladium catalyst and DPPF ligand. Those reactions in

which the zinc cyanide was the limiting reagent, either failed or produced a yield of between 8 and 13% over the time scale of 48 – 72 hours. For these reactions dimethylacetamide was the solvent selected. Further literature indicated that the use of zinc acetate along with zinc cyanide and zinc powder prolonged the catalytic lifespan of the active palladium species and contradictory to previous literature, water was used alongside the organic solvent DMF.[285] However, such procedures under reaction times of 24 – 96 hours, resulted in low yields of 12%, or failure to produce the desired compound at all. Success was achieved using molar excess of zinc cyanide and zinc powder, $Pd_2(dba)_3$ and DPPF in anhydrous dimethylacetamide. The reaction was heated to 120 ˚C for 1 hour yielding the desired nitrile **3.16** (Scheme 3.1) in 85% yield.

Initial attempts to reduce the nitrile to its corresponding primary amine **3.17** (Scheme 3.1), involved refluxing the nitrile with the ligand $NiCl_2.6H_2O$ and the reducing agent sodium borohydride in methanol for 12 hours; however no product was isolated.[286] Subsequent reaction of the nitrile with aqueous ammonia solution and Raney nickel under a hydrogen atmosphere, in a mixture of tetrahydrofuran and methanol at 25 ˚C for 3 hours,[287] gave the crude product which was confirmed by mass spectrometry, ^{1}H and ^{13}C NMR. The crude product was used without further purification in the following reaction, protection of the primary amine.

When selecting the amino protecting group a number of considerations were taken into account, including the need for selective de-protection of the *t*-butyl ester without any effect upon the amine functionality. This meant that the group introduced needed to be stable under acidic conditions and base labile. The trifluoroacetamide group which at room temperature is stable under acidic conditions and can be removed under basic conditions (including pH 12), was selected as a suitable candidate.[279] Incorporation of the trifluoroacetamide group was successful, affording the desired compound **3.18** (Scheme 3.1) in a 68% yield. With the amino functionality now protected, the *t*-butyl protecting group could then be removed, un-masking the carboxylic acid **3.19** for the coupling reaction with the selected solubilising groups.*t*-Butyl ester cleavage was carried out using a 1:1 mixture of trifluoroacetic acid and DCM at 25 ˚C, giving full conversion to the TFA salt of the carboxylic acid **3.19** (Scheme 3.1) after 3 hours. As with the amine **3.17** (Scheme 3.1) the TFA salt of the carboxylic acid **3.19** was identified using mass spectrometry, ^{1}H and ^{13}C NMR and carried forward to the subsequent reaction without need for purification.

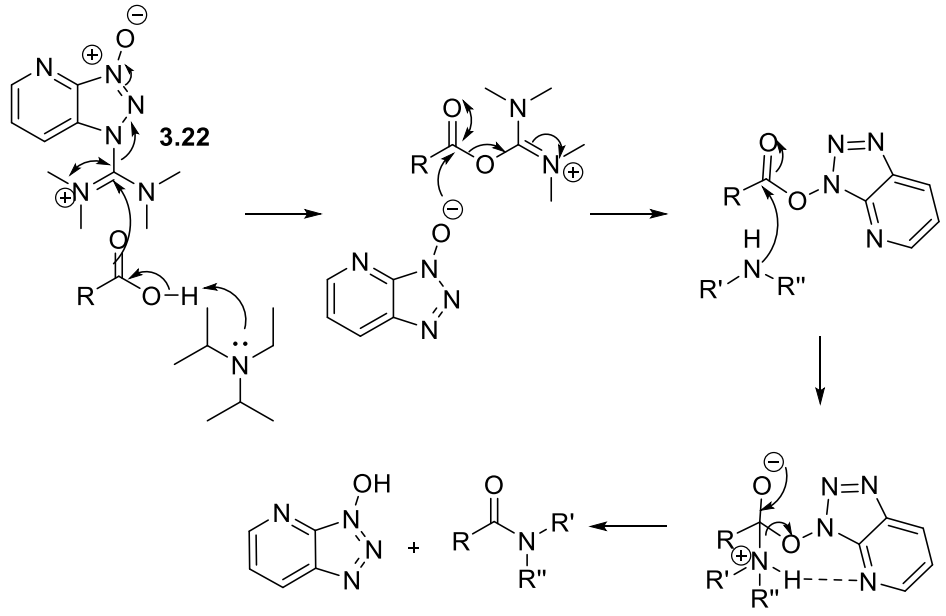
The outcome from the computational modelling indicated that the top 10 selected targets for synthesis contained *N*-methylpiperazine and morpholine derived amides or amines. Peptide coupling reagents were used to introduce these groups using the crude carboxylic acid **3.19**, the initial targets being the *N*-methylpiperazine derived amides. Formation of the amide bond between the carboxylic acid **3.19** and the *N*-methylpiperazine, was initially investigated using the coupling reagent EDCI. Coupling using EDCI and DMAP was unsuccessful with no product isolated. Consequently an alternative coupling reagent was used. 1-[Bis (dimethylamino) methylene]-1H-1,2,3-triazolo [4,5-b] pyridinium 3-oxihexafluorophosphate (HATU) belongs to another family of reagents associated with 1*H*-benzotriazole.[250] There are two known forms of the reagents, the uronium species (or *O*-form **3.22**) and the guanidinium species (*N*-form **3.23**), which when in solution are believed to be in an equilibrium with one another (Scheme 3.3).[249,250]



O-form **3.22** ⇌ *N*-form **3.23**

Scheme 3.3- Equilibrium between the *O*-form and *N*-form of HATU adapted from [249].

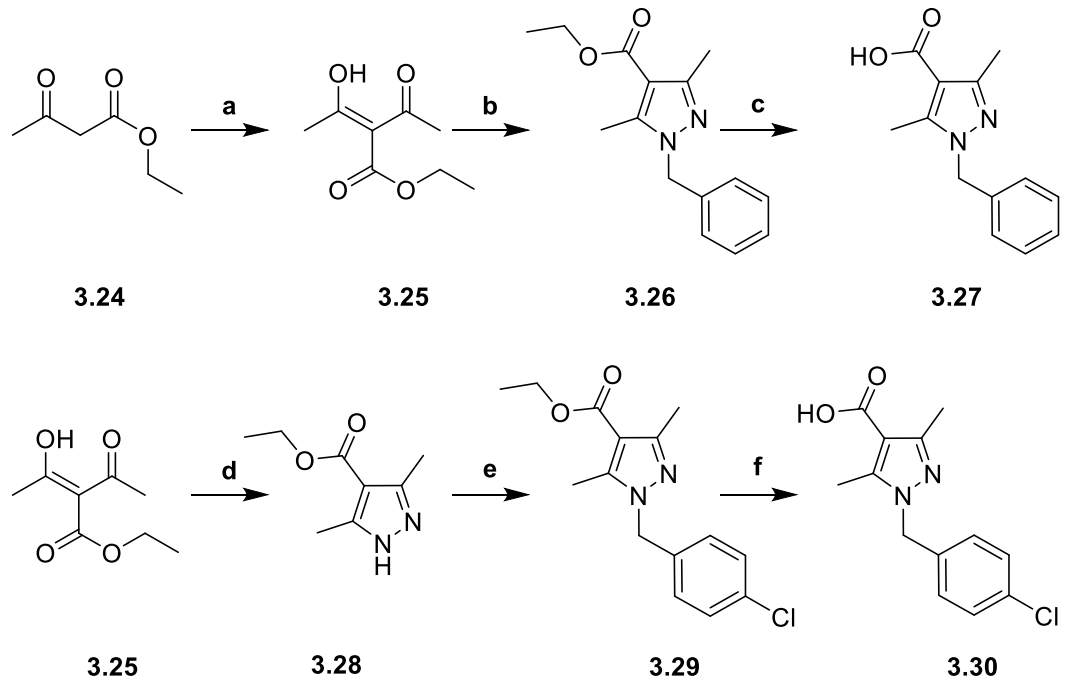
The reduction of racemisation and efficiency of HATU with more sterically hindered reactants, make it a popular reagent for use in both solution and solid-phase peptide synthesis.[288] The mechanism of the reaction involves activation of the carboxylic acid by HATU followed by *N*-acylation which can be seen in Scheme 3.4. Initially the carboxylic acid is deprotonated by the organic base, thus forming the carboxylate anion which goes on to attack the *O*-form HATU **3.22** to form an unstable salt. This unstable intermediate is rapidly converted to the activated ester of 7-aza-HOBt which is an extremely reactive species towards amines. Finally, nucleophilic addition of the amine, results in *N*-acylation, yielding the desired amide product.[249,289]



Scheme 3.4- Proposed reaction mechanism of HATU **(3.22)** couplings with generic amines.

With this in mind, a general procedure for the coupling of *N*-methylpiperazine to the acid **3.19** was explored; this involved dissolving the acid in anhydrous DMF followed by the addition of a molar excess of HATU, DIPEA and finally *N*-methylpiperazine. The reaction was heated to 40 ˚C for 48 hours but once again, as with the EDCI reaction, proved unsuccessful as no product was isolated upon workup.

The coupling reactions for the formation of morpholine derived amides were also investigated using the HATU procedure. Monitoring of the reaction using TLC indicated that what was believed to be the desired product was indeed forming, although slowly. Despite initial attempts being low yielding, optimisation of the reaction was successful and isolation of the product after a total reaction time of 96 hours at 25 – 40 ˚C afforded the desired amide **3.20** (Scheme 3.1) in a 55% yield. The resultant deprotection of the trifluoro acetylated amine to compound **3.21** (Scheme 3.1) was carried out by stirring the precursor in 7N methanolic ammonia over 96 hours, generating the product in 31% yield. The starting material was recovered and re-subjected to the reaction conditions to maximise the yield for the subsequent amide coupling with compounds **3.27** and **3.30**.



3.24 **3.25** **3.26** **3.27**

3.25 **3.28** **3.29** **3.30**

Scheme 3.5- Synthetic route to fragments 3.27 and 3.30. (a) acetyl chloride, chloroform, Mg turnings, EtOH, toluene,119 ˚C, 24 hrs, 59%. **(b)** benzyl hydrazine, EtOH, 85 ˚C, 51%. **(c)** 1M NaOH, MeOH, 68 ˚C 1 hr, 25 °C 24 hr, 89%. **(d)** Hydrazine, EtOH, 25 °C, 16 hrs, 93%. **(e)** 4-chlorobenzylbromide**,** cyclohexanone, K_2CO_3, 155 ˚C, 51%. **(f)** 1M NaOH, MeOH, 68 ˚C 1hr, 25 °C 24 hrs, 34%.

The synthetic route for the fragments **3.27** and **3.30** (Scheme 3.5) involved acylation of ethyl acetoacetate to ethyl 2-acetyl-3-hydroxybut-2-enoate **3.25**, followed by a cyclo-condensation reaction to form the pyrazole ring of **3.26**. Subsequent ester hydrolysis afforded the fragment **3.27** with an 89% yield.[290] The second fragment, **3.30** (Scheme 3.5) was also formed by an initial cyclo-condensation of ethyl 2-acetyl-3-hydroxybut-2-enoate **3.25** to 3, 5–dimethyl-1*H*-pyrazole-4-carboxylate **3.28**. Benzylation with 4-chlorobenzylbromide of the pyrazole ester **3.28** to compound **3.29**, was followed by ester hydrolysis to afford product fragment **3.30** in a 92% yield. Compounds **3.24 – 3.30** were synthesised by 4th year MChem student Solon Mardapittas following procedures previously devised for his project.

With all three major fragments synthesised the final step in the synthesis of two of the analogues of Inhibitor 2 was the amide coupling reaction between the amine **3.21** and the two carboxylic acids **3.27** and **3.30** respectively. The procedure employed was the HATU coupling reaction used previously, with substitution of the base DIPEA for potassium carbonate (K_2CO_3). In both cases the reaction was successful in yielding the desired products Inhibitor 4 and Inhibitor 5 with yields of 36% and 34% respectively. The two Inhibitors were then tested *in vitro* and the data can be found in the following Section 3.3.

3.2.3 Summary

The initial aim of the second generation chemical synthesis was to synthesise a number of small molecule compounds, ligands which shared a common core scaffold with Inhibitor 2.

The initial molecular target **3.6** was selected due to the prevalence of the *N*-methyl piperazine functional group (Figure 3.6) present within four of the 12 compounds. The initial five steps of the synthetic route could be applicable to 7 of the desired 12 compounds with possible modifications needed for those ligands containing pyridone moieties. The numerous steps to reaching the carboxylic acid **3.19** were not without difficulties, as such the optimisation of these procedures required more time than expected. The subsequent step involved coupling of the acid **3.19** to the different solubilising moieties, to produce the desired amide, for example **3.20**, which could then be selectively reduced to the respective amine if required. Initial efforts to couple *N*-methylpiperazine to **3.19** were unsuccessful. As four of the top 12 compounds were morpholine analogues, this was the most logical next group to couple. The HATU coupling procedure employed to access the desired amide **3.20** was successful with a pleasing yield of 55%.

During the synthesis of compounds **3.14 – 3.21,** 4th year MChem student Solon Mardapittas was working in parallel to synthesise the fragments **3.27** and **3.30.** These two fragments would be coupled with analogues such as **3.20,** to achieve the desired amide final compound using the HATU procedure previously developed. Restrictions in time, associated with the optimization of the synthetic route, meant that no further analogues were developed. Instead Inhibitors 4 and 5 were synthesised with yields of 36% and 34% respectively and carried forward for biophysical analysis.

3.3 Binding assessment using Biophysical Methods

With the second generation of inhibitors synthesised, biophysical assessments were made of the ligands interactions with NCS1. To achieve this the experimental techniques used to analyse the first generation (Chapter 2.3), were again employed with these second generation molecules. For further details on the techniques used please see Chapter 1.6 and Chapter 2 sections 2.3.3, 2.3.5 and 2.3.7.

3.3.1 NMR Spectroscopy

General experimental procedures

The analysis techniques and experimental procedures, including the conditions used to acquire the NMR spectroscopy screening data of the small molecules interactions with NCS1, can be found in Chapter 6.3.2.7. However, briefly for NMR investigation all NMR spectra were collected at 298K, on either a Bruker AVANCE II+ 600 MHz or 800 MHz spectrometer, equipped with 5 mm TCI triple-resonance cryoprobes.

All protein samples were prepared in the NMR buffer and were of the concentration 50 μM, full details on the preparation of NCS1 can be found in Chapter 6.3.2. The NMR data collected for investigation included 2-D ^{1}H ^{15}N HSQC, ^{1}H ^{13}C TOCSY. The 2-D HSQC binding screens were carried out at a protein: inhibitor ratio of 1: 10 and 1: 20 whereby the protein concentration was 50 μM and the inhibitors 500 μM and 1000 μM respectively. For the purposes of the discussion of the changes in the aliphatic and aromatic methyl spectra, the methyl groups of the amino acid residues are labeled in single letter nomenclature a=alpha, b=beta.

Inhibitors were kept in stock solutions at 100 mM (m/v) concentration in 100% DMSO and were added (v/v) to the NMR samples. The chemical shift perturbations caused upon an interaction of the small molecules with NCS1 are referenced against the interactions of NCS1 with the same amount (v/v) of DMSO, which is at the same percentage as in the inhibitor sample and never exceeded 5% (v/v).

3.3.1.1 Solubility

The reduced solubility of the first generation of inhibitors in aqueous buffer, prevented accurate quantitative assessment from the biophysical techniques. It was found to be particularly detrimental to the determination of the binding affinity as discussed previously (Chapter 2 sections 2.3.5 and 2.3.6).

The second generation of inhibitors were designed with the addition of a solubilising group to overcome this problem (3.1.2). A simple 1-D experiment was conducted which involved making an NMR sample of each inhibitor at 500 μM (m/v) in NMR buffer with a reference solvent trimethylsilyl propanoic acid (TSP) at 100 μM (Figure 3.7).

The TSP reference peak can be seen at 0 ppm and the integral height for this peak is set to a standard value between each spectra. From this data the concentrations in solution of inhibitors 1, 2 and 3 were significantly less than inhibitors 4 and 5.

This is apparent in the aromatic region (Figure 3.7 c) and is indicative that one of the primary aims of the second generation of inhibitors “to improve the solubility” was achieved.

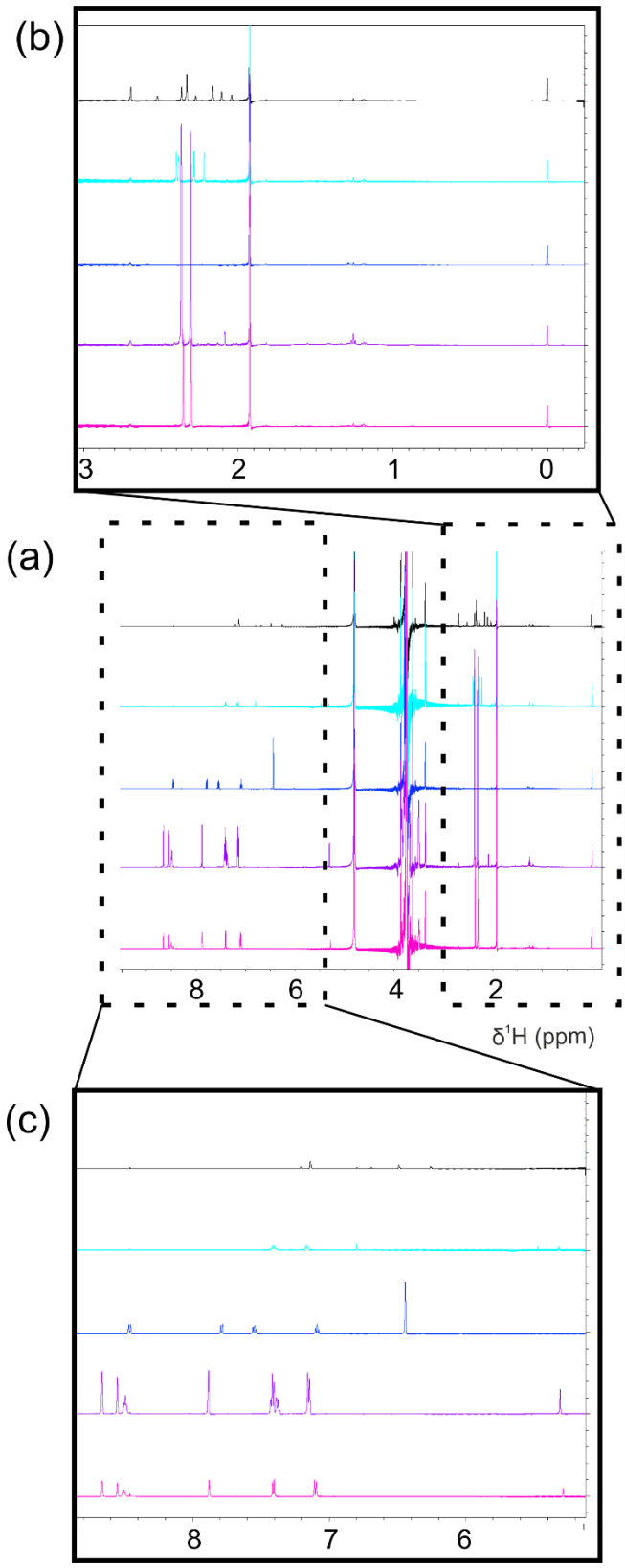


Figure 3.7- **Solubility screen of inhibitors 1 - 5.** Inhibitor 1 (black), Inhibitor 2 (cyan), Inhibitor 3 (blue), Inhibitor 4 (purple) and Inhibitor 5 (magenta). Concentration of all inhibitors 500 µM all samples prepared from separate 100 mM stock solutions made up in 100% DMSO added into NMR buffer pH 6.4 with 100 µM TSP acquired on Bruker AVANCE II 600 MHz spectrometer at 298K total percentage of DMSO in each sample was 0.5%. **(a)** Aliphatic region enlargement. **(c)** Aromatic region enlargement. The solubility of the second generation of inhibitors (4 and 5) can be seen as being greatly improved in comparison to those of the first generation (1, 2 and 3).

3.3.1.2 Ligand interactions by NMR spectroscopy

With the improved solubility of the ligands, the next stage was characterisation of ligand interactions with the protein NCS1. This was conducted at two protein: ligand ratios, 1:10 and 1:20, with their improved solubility the concentration of both inhibitors was achievable at a substantial excess of the protein.

Despite the improved solubility of the second generation of inhibitors, they were still not fully water soluble at the desired concentrations and so were solubilised in DMSO for the purposes of biophysical characterisation. As with the first generation of ligands it was imperative to carry out identical 2-D ^{1}H ^{15}N HSQC control experiments with the same concentration of DMSO as would be present in the inhibitor containing samples (Figure 3.9). In doing so we were able to determine if the changes in the chemical shifts of the amino acid residues were true to the Inhibitor, or were an artefact of an interaction with DMSO.

Inhibitors 4 and 5 are very similar in their chemical structure and only differ with the addition of a para-chloro group on the phenyl ring of Inhibitor 5 (Figure 3.8). The additional 4-chloro/ para-chloro group was selected following the Topliss scheme for optimisation of phenyl ring substitution (discussed further in Section 3.1.2).[273]

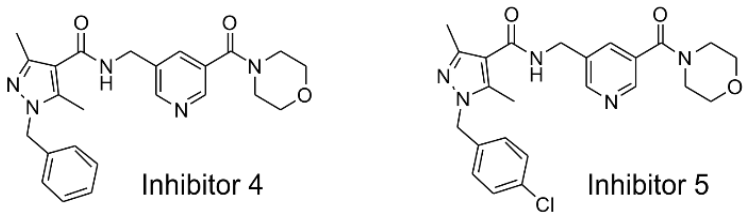


Figure 3.8- The chemical structures of the second generation of designed and synthesised small molecules, Inhibitor 4 and Inhibitor 5, both analogues of Inhibitor 2 (Figure 3.1).

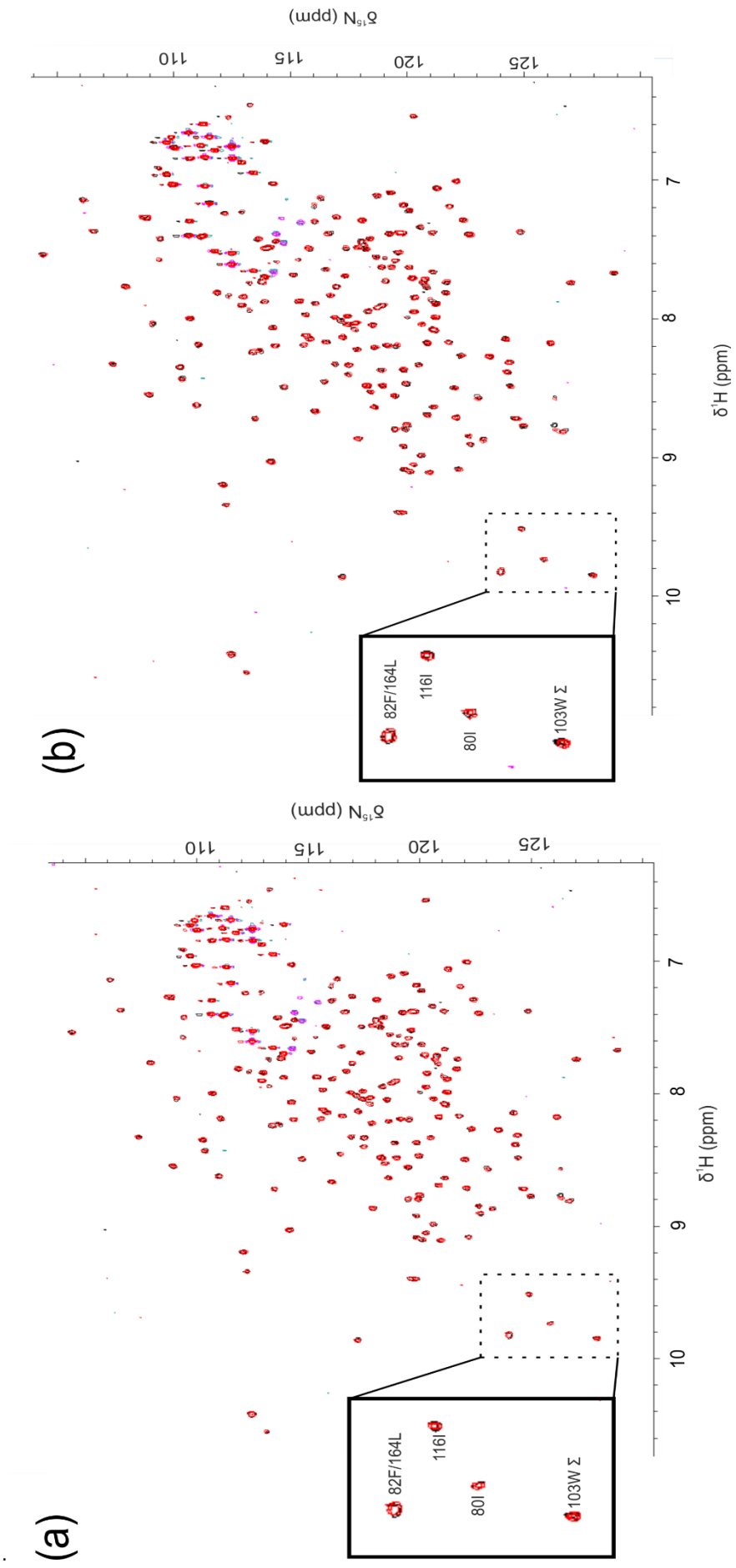


Figure 3.9 - NCS1 interactions with DMSO, control spectra. Two 2-D ^{1}H ^{15}N HSQC spectra of ^{1}H ^{15}N NCS1 50 µM acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 800 MHz spectrometer. **(a)** ^{1}H ^{15}N NCS1 50 µM black overlaid with ^{1}H ^{15}N NCS1 50 µM and 0.5% DMSO red. No changes in the spectrum indicating that at a concentration of 0.5% DMSO does not interact with NCS1. **(b)** ^{1}H ^{15}N NCS1 50 µM black overlaid with ^{1}H ^{15}N NCS1 50 µM and 1.0% DMSO red. One minor peak change indicates a very small interaction between NCS1 and DMSO at 1%.

Despite the almost identical chemical structures, it was apparent that Inhibitor 4 appeared to exert a lesser effect on NCS1 than Inhibitor 5. This can be inferred from analysis of the 2-D ^{1}H ^{15}N HSQC spectra (Figure 3.10), where there are fewer peak attenuations and the changes in chemical shift of the residues when inhibitor 4 is present. As previously discussed in Chapter 2; calculation of the "chemical shift perturbation" (Chapter 2 Section 2.3 eqn 2.1), allows for the determination of residues within the protein that are effected by a ligand. As previously defined the sum of the total number of residues perturbed was calculated and from this value the top 10% and 20% ranges were determined. The residues whose chemical shift perturbation sum fell within these two ranges were defined as being within the top 10 and top 20 percentile respectively.

When comparing the values of the changes in chemical shift, Inhibitor 4 caused an average change of between 0.00043 and 0.0405 at 500 μM (P:L 1:10), which subsequently increased to between 0.0043 and 0.083 as the concentration of ligand was increased to 1 mM (P:L 1:20). This indicated that saturation of the protein by the inhibitor had not been achieved even at 500 μM and also suggested that the binding affinity of Inhibitor 4 could be relatively weak as the ligand was now in a 20 fold molar excess to that of the protein.

Further determination of the structural effects of Inhibitor 4 on NCS1 through interpretation of NMR chemical shift perturbation data, indicated that at a concentration of 500 μM the inhibitor induced attenuation of a number of residues. When compared to the DMSO control, in which no attenuation occurred it can be concluded that these attenuations occurred directly as a result of Inhibitor 4; affected resonances are from residues: Asp37, Tyr52, Ile179, Gln181 and Leu183.

Furthermore, those residues whose shifts, rather than linewidths, are greatly affected by the presence of the inhibitor, are within the top 10 and 20 percentile group of residues that experience chemical shift effects. By combining the attenuated residues with those found within the top 10 and 20 percentile ranges, allowed the regions of NCS1 that interact with the ligand to be identified.

In the case of Inhibitor 4 the residues whose resonances are affected by its presence include residues from the hydrophobic binding groove (Figure 3.13). The regions implicated in the interaction with the *C*-terminal helix of the D2R peptide are affected by the presence of Inhibitor 4.[38] Interestingly of the residues affected three Asp37, Tyr52 and Val68 are the original key points of contact targeted in the initial design process of this project (Chapter 2.1.2 Figure 2.2), they are also found to be present

in the top binding pose of Inhibitor 4 determined through the GOLD docking analysis (Figure 3.13). Increasing the concentration of Inhibitor 4 to 1 mM led to further attenuation of some residues including Val13, Leu101, Ala104 and Ile128 (Figures 3.11 and 3.13). The position of these residues indicates that increasing the concentration of the ligand, increases the non-specific binding tendencies, with further points of contact away from the desired hydrophobic face.

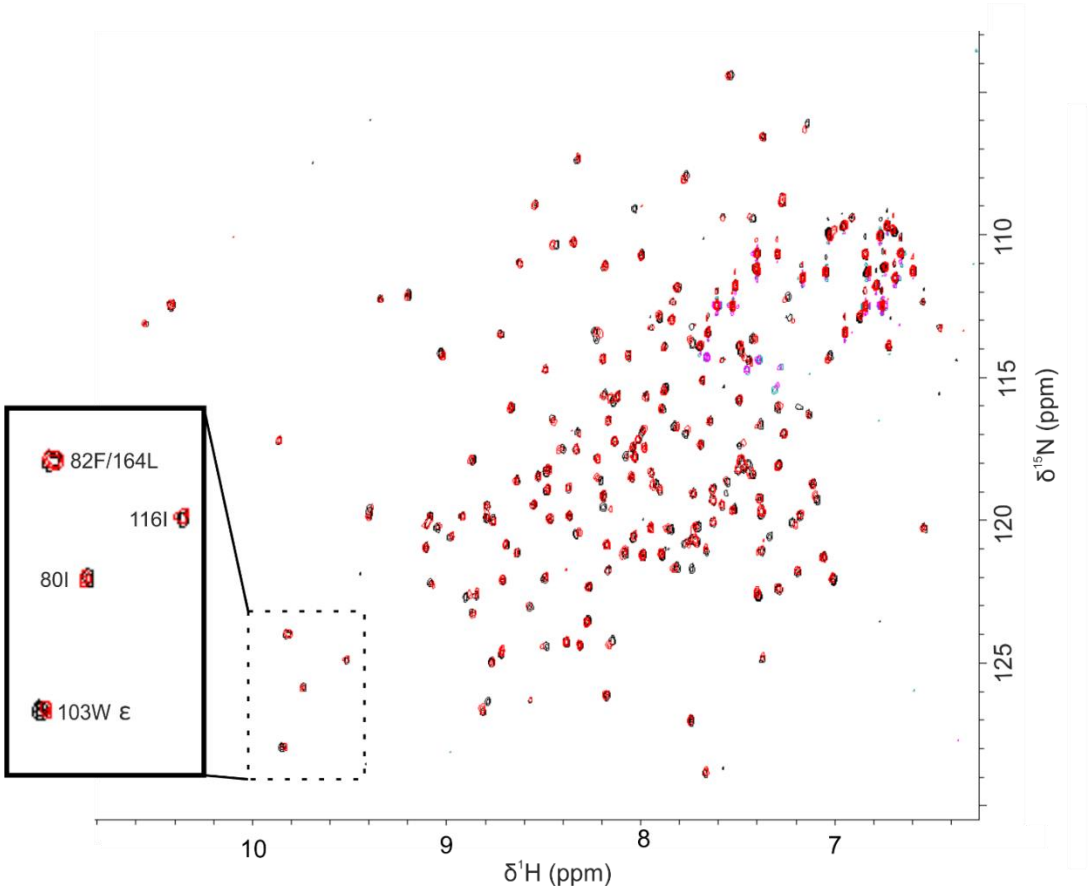


Figure 3.10- **Ligand binding screen NCS1 with Inhibitor 4.** A 2-D ^{1}H ^{15}N HSQC spectra of ^{1}H ^{15}N NCS1 50 µM and 0.5% DMSO black overlaid with ^{1}H ^{15}N NCS1 50 µM and Inhibitor 4 500 µM P:L ratio of 1:10, ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 800 MHz spectrometer at 298K. . Changes in chemical shift of some peaks such as Trp103 are indicative of an interaction between the inhibitor and the protein. The residues highlighted within the box were selected to indicate changes in specific residues as examples.

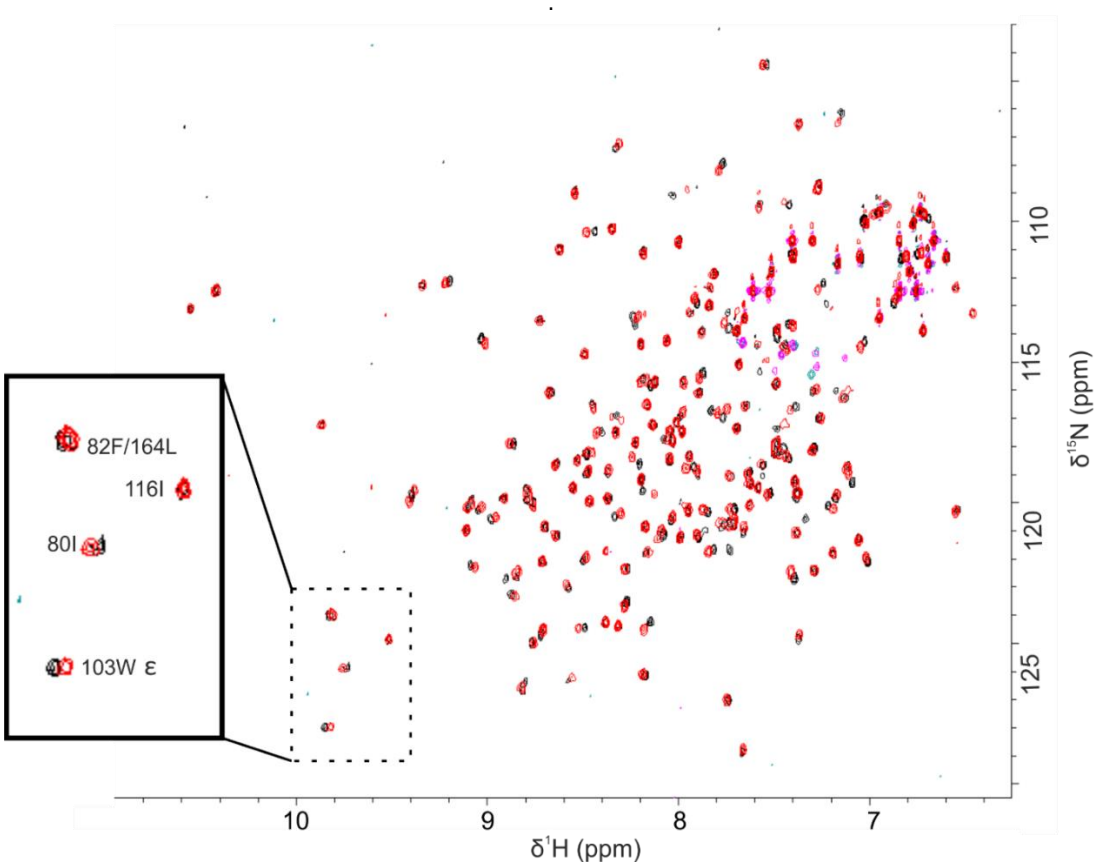


Figure 3.11- **Ligand binding screen NCS1 with Inhibitor 4.** A 2-D ^{1}H ^{15}N HSQC spectra of ^{1}H ^{15}N NCS1 50 µM and 1.0% DMSO black overlaid with ^{1}H ^{15}N NCS1 50 µM and Inhibitor 4 1000 µM P:L ratio of 1:20. ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 800 MHz spectrometer at 298K. Increasing the concentration of Inhibitor 4 causes further changes in the chemical shift of some residues including Trp103 and Ile80 indicating the protein had not reached saturation at 500 µM. The residues highlighted within the box were selected to indicate changes in specific residues as examples.

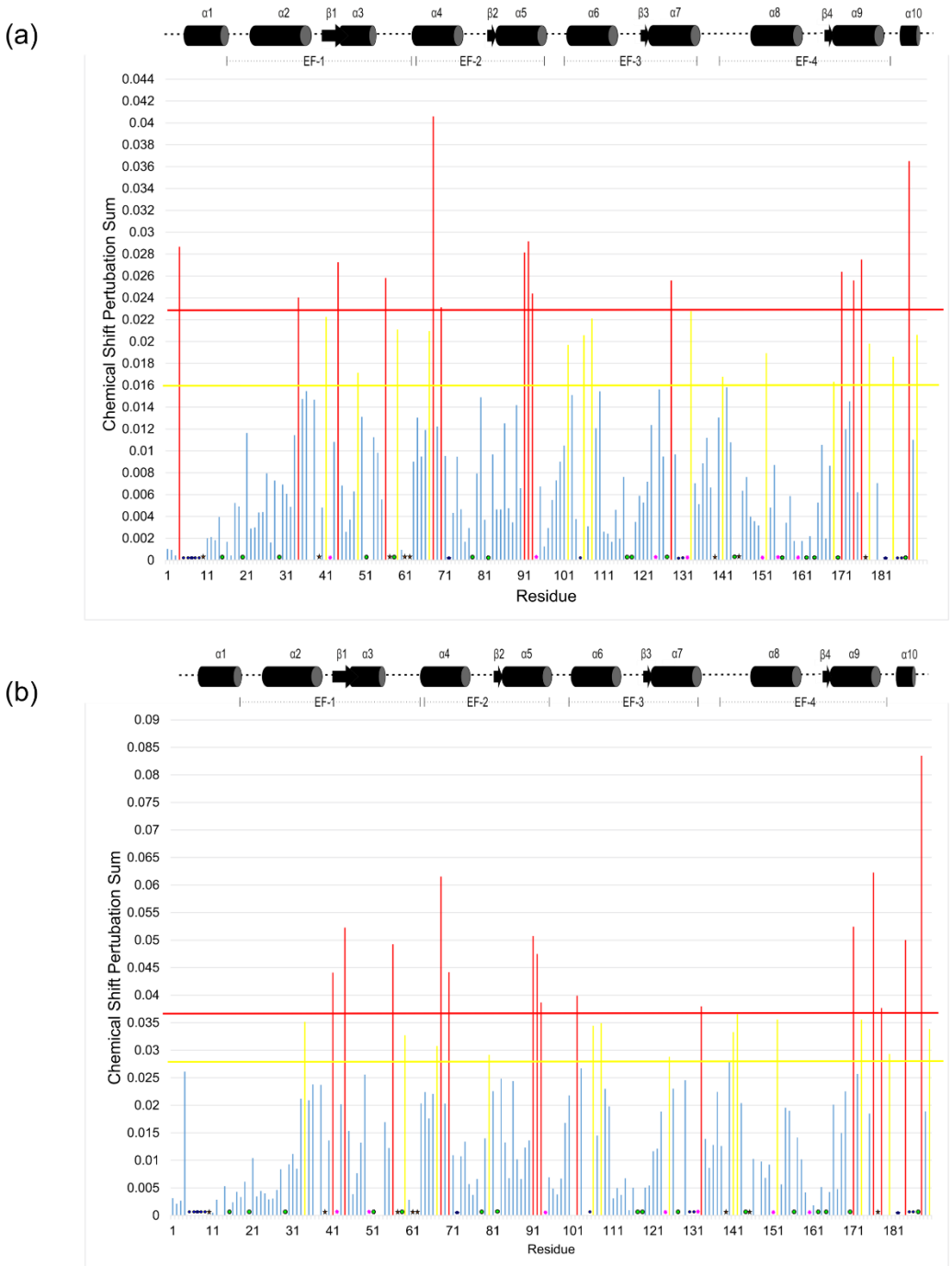


Figure 3.12– (a) Histogram of the perturbations of NCS1 residues from ligand binding screen NCS1: Inhibitor 4 at 500μM P:L 1:10. The changes in chemical shift pertubation for each residue was calculated using equation 2.1; the changes are classified as percentile values of the total number of residues affected for analysis. Those residues whose chemical shift underwent a change in ppm in the top 10 percentile can be seen above the red line and these include: Val68, Gly188, Thr92, Lys4, Val91, Asp176, Asp44, Glu171, Phe56, Lys174, Ile128, Ser93, Phe34, Asn70. The yellow line represents those residues who's change in chemical shift was within the top 20 percentile and includes Gly133, Gly41, Tyr108, Gly59, Phe67, Val190, Lys106, Ser178, Arg102, Ile152, Ser184, Gln49, Glu141, Phe169. **(b)** Histogram of the perturbations in NCS1 residues from ligand binding screen NCS1: Inhibitor 4 at 1 mM P:L 1:20, those residues whose chemical shifts underwent a shift in the top 10 percentile can be seen above the red line and these include: Gly188, Asp17, Val68, Glu171, Asp44, Val91, Ser184, Phe56, Thr92, Asn70, Gly41, Arg102, Ser93, Gly133. The yellow line represents those residues whose chemical shift change were within the top 20 percentile and includes: Ser178, Glu142, Ile152 Ser173S, Phe34F, Tyr108, Lys106, Val190, Glu141, Gly59, Phe67, Val180, Ile80, Val125. Unassigned residues Single peaks assigned with two residues Proline peaks that became too broad to assign.

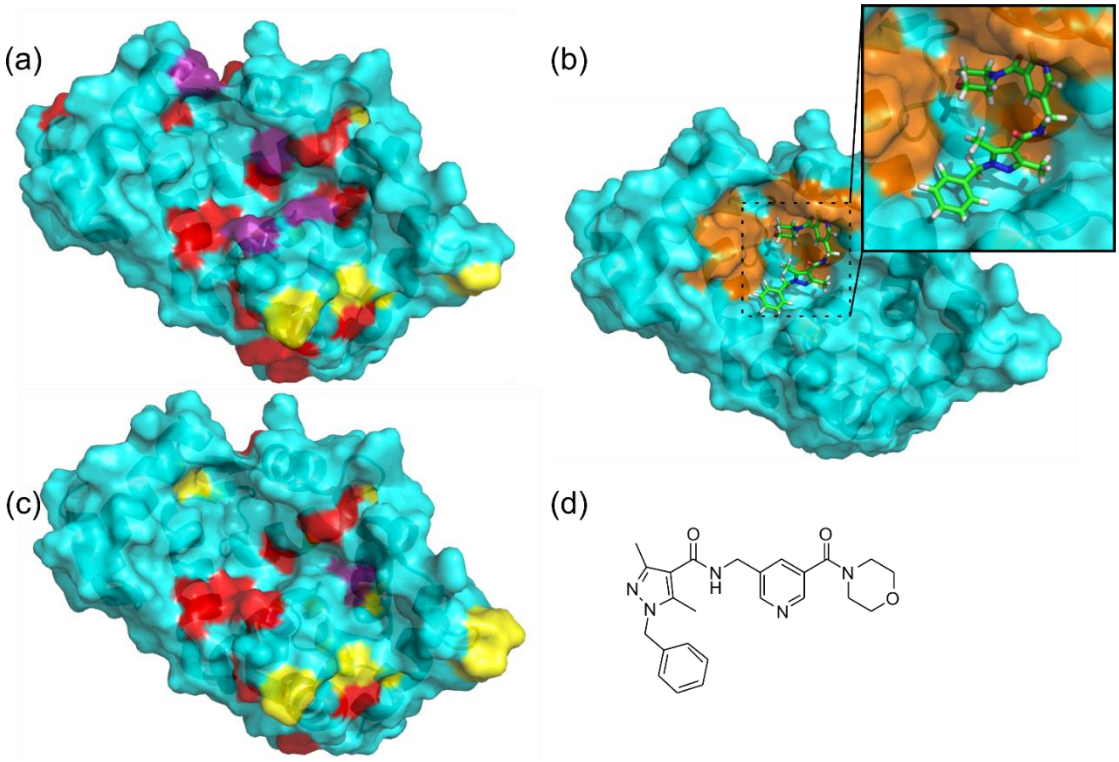


Figure 3.13 – (a) Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 4 at 500 μM determined from the corresponding histogram, the residues within the top 10 percentile are coloured in red and top 20 percentile in yellow, those coloured in purple belong to the residues that were attenuated upon the addition of the ligand and not due to the affects of DMSO (PDB 5AER). **(b)** Prediction of the binding pose for the interaction between Inhibitor 4 and NCS1 (3.1.2) those key residues highlighted in orange.[4] **(c)** Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 4 at 1 mM determined from the corresponding histogram the residues within the top 10 percentile in red and top 20 percentile in yellow, those coloured in purple belong to the residues that were attenuated upon the addition of the ligand and not due to the affects of DMSO (PDB 5AER). **(d)** Chemical structure of Inhibitor 4.

As with the biophysical characterisation of the first generation of inhibitors, calculation of the chemical shift pertubation for the backbone residues was followed by analysis of the effects of Inhibitor 4 on the aliphatic methyl and aromatic sidechain residues. To perform this analysis of the difference between the control ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 μM and 0.5% DMSO with the ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 μM and Inhibitor 4 500 μM (Figure 3.14) were evaluated.

Inhibitor 4 appears to effect seven aliphatic methyl sidechain residues, of which only one residue is completely attenuated; Thr144 located on the loop of EF-4. It does however, cause partial attenuation of 7 residues in that one of a pair of methyl groups are effected: Leu_a16, Val_a71, Val_b132, Ile_a152, Val_a180 and Val_b190 As well as attenuations, Inhibitor 4 appears to also cause a number of notable changes in chemical shift including Ile51, Ile80 and Leu89 located on EF-1 and EF-2. As with previous inhibitors these residues are not located in a specific region of NCS1 but are in fact found across the four EF hand regions on α helices and the terminal flexible loop.

With the aromatic residues, as with the first generation ligand Inhibitor 3, the Tyrε residues are unaffected by Inhibitor 4 (Figure 3.15); however residues that are attenuated include: Phe22, Phe67, Phe85, Phe153 and Phe169 primarily located in α helical regions of EF-2 and EF-4.

Analysis of the effects of Inhibitor 4 on the sidechain residues is in agreement with the 2-D ^{1}H ^{15}N HSQC spectra, in that the observed change in the presence of the inhibitor is more subtle.

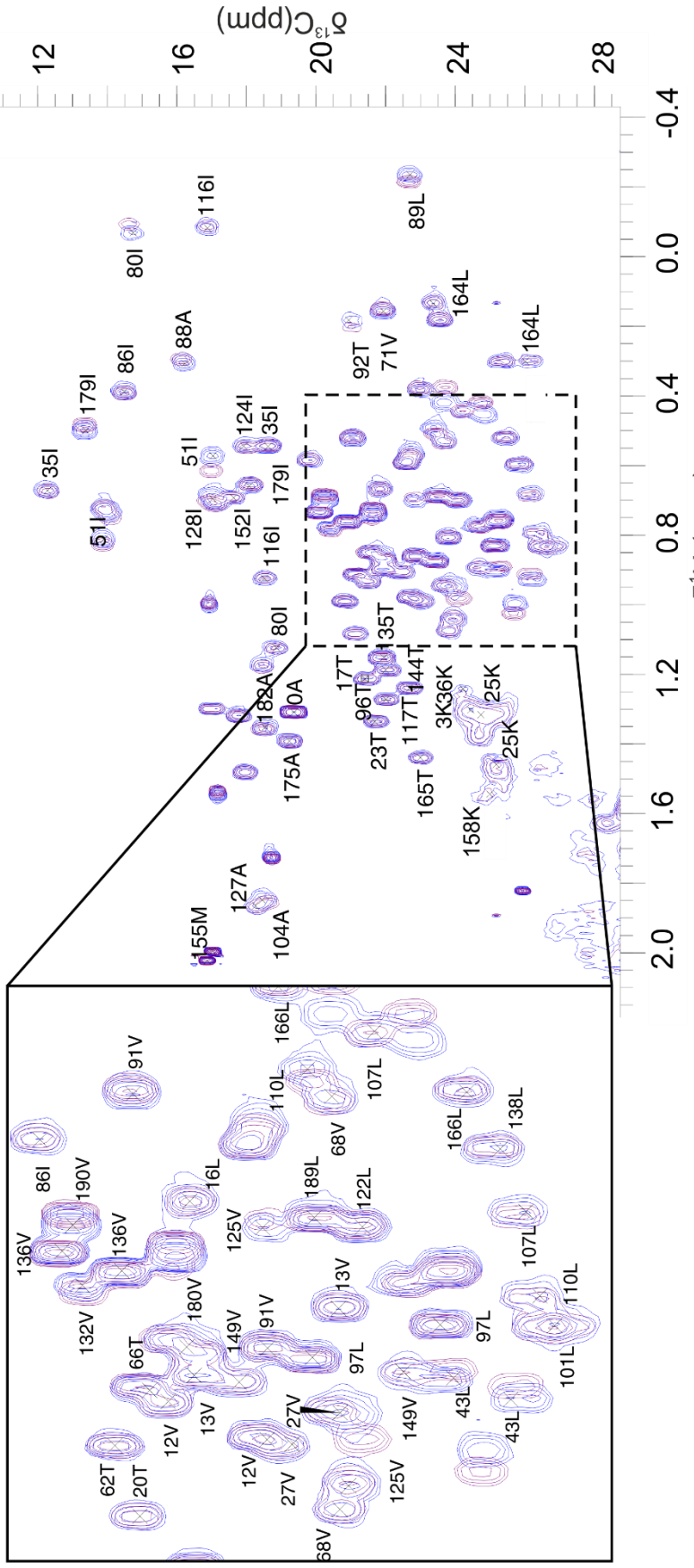


Figure 3.14- An overlay of two aliphatic methyl sidechain regions of NCS1 with and without Inhibitor 4. The ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 0.5% DMSO (purple) overlaid with the ^{1}H ^{15}N ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 4 500 µM (blue). Those residues of NCS1 related to the Inhibitor 4 spectra for which the transfer of assignment was possible can be seen, there a number of notable changes in chemical shift including Ile51, Ile80 and Leu89. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

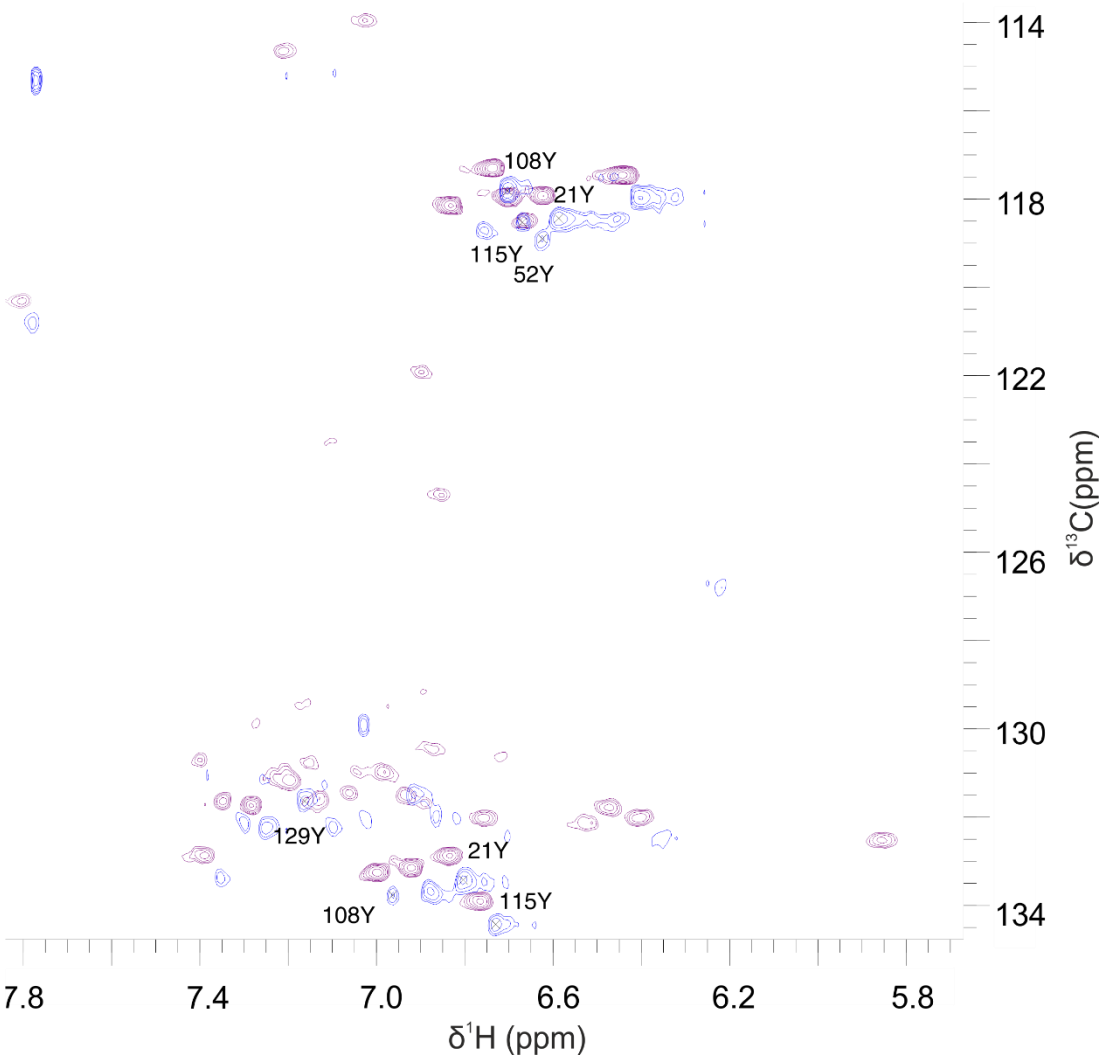


Figure 3.15- Two overlaid aromatic sidechain spectra of NCS1 with and without Inhibitor 4. The aromatic ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 μM and 0.5% DMSO (purple) overlaid with the ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 μM and Inhibitor 4 500 μM (blue), the aromatic residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same peak, or those which were attenuated by the presence of Inhibitor 4 including; Phe22, Phe67, Phe85, Phe153 and Phe169. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

Further to the investigation of the interaction between Inhibitor 4 and NCS1 in relation to the positions on the protein that are effected by the ligand, we were also able to determine the chemical moieties of the ligand that were implicated in the interaction with NCS1.

To achieve this we employed two experiments: 1-D ^{1}H saturation transfer difference (STD) and 1-D ^{1}H water ligand observation with gradient spectroscopy (waterLOGSY). The experimental processes shall not be discussed in detail in this chapter, more information can be found within Chapter 4. However in short, both experiments work on the basis of the transfer of magnetisation, the difference between the two lies in where the saturation of magnetisation is initiated. In the case of the STD experiment this is a resonance of the protein spectra where no ligand

signals can be found usually upfield of 0 ppm.[291] With the waterLOGSY this position is the bulk water that can be found within binding regions between proteins and ligands.[292][293] In the STD experiment, if there is a binding event, magnetisation is transferred from the irradiated protein to the bound ligand which is in a dynamic equilibrium with the free ligand in solution resulting in the appearance of the ligand signals within the difference spectra. This process was conducted for inhibitors 1, 2 and 3; however due to their limited solubility, no meaningful data was obtained using this technique; example spectra can be found in the appendix (Section A.1).

The STD and waterLOGSY spectra of Inhibitor 4 500 µM with NCS1 50 µM (Figure 3.16 b) was compared to the same spectra of the Inhibitor alone in NMR buffer and at the same concentration; the presence of ligand signals in the difference spectra in the complex is indicative of a binding interaction. This is observed when we directly compare the two STD spectra (Figure 3.16 a and b (black spectral line)), in the control STD experiment there is no evolution of the ligand signal as there is no protein present to be irradiated and hence no possible binding interaction. In the presence of NCS1 we are able to see the hydrogen signals from the following atomic positions: 1, 2, 3, 4, 5, 6, 7, 9, 10 and 11 on the inhibitor (Figure 3.16 b). This is in concordance with the magnetisation being transferred from the protein, to the bound ligand which is in an exchange equilibrium with free ligand in solution and hence detection of the ligand signals.

Analysis of the two waterLOGSY spectra showed a reduction in the intensity of the ligand signals of Inhibitor 4 between the buffer control and the protein containing sample, indicative of an interaction between the protein and the Inhibitor. The NH group 7 appears positive as it is exchangeable in nature. Hence, both the STD and waterLOGSY experiment indicate that there is a binding event occurring between Inhibitor 4 and NCS1, although the relative affinity cannot be easily determined.

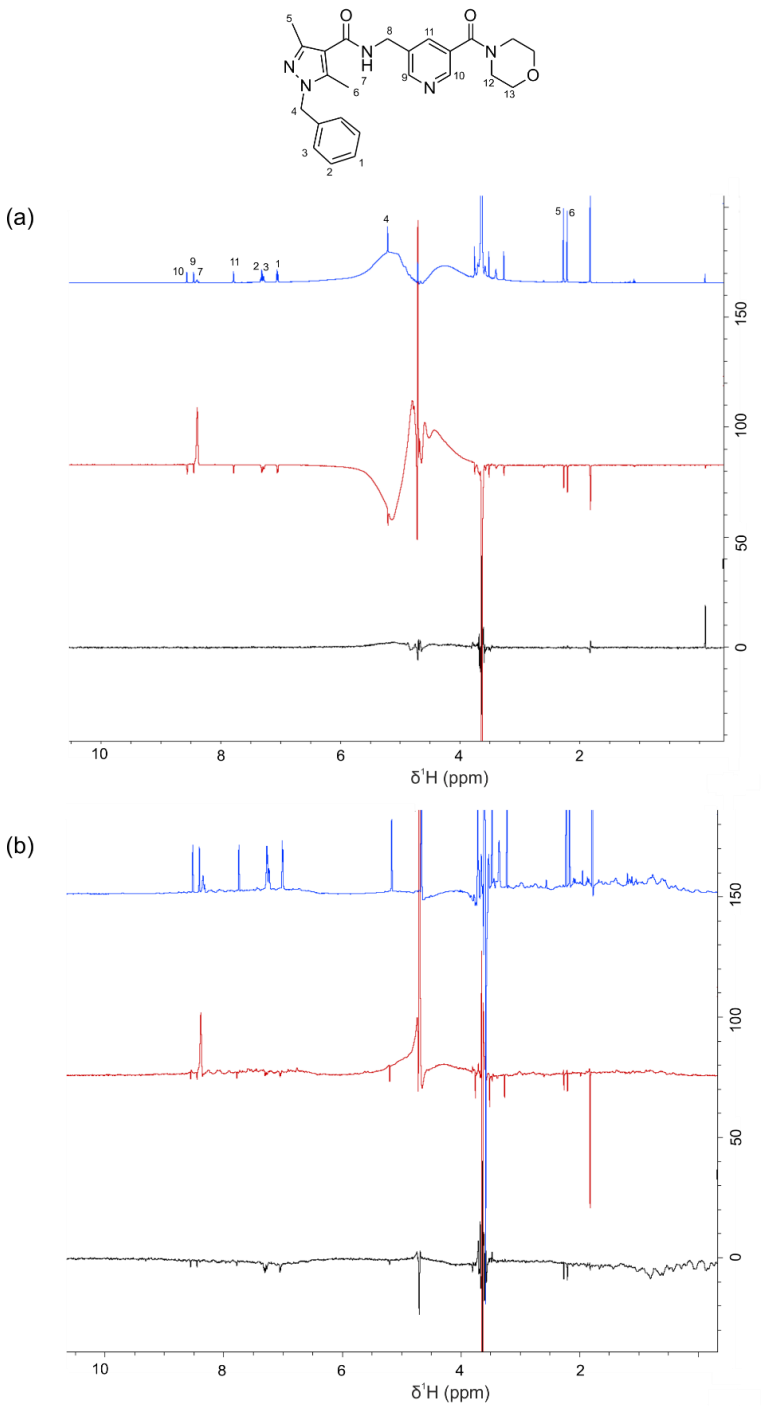


Figure 3.16- 1-D ligand screening spectra of Inhibitor 4. (a) Buffer standard 500 µM Inhibitor 4 in NMR buffer; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(b)** Binding screen 500 µM Inhibitor 4 and 50 µM NCS1 in NMR buffer; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). Comparison between the buffer controls (a) and the protein containing experiments (b) indicate possible binding events occurring at ligand positions 1, 2, 3, 4, 5, 6, 7, 9, 10 and 11. This is observable in the STD spectra where the relative ligand signals are apparent and negative in the waterLOGSY spectra the relative intensities of the peaks are reduced slightly. Resonances 8, 12 and 13 are not observed as they appear in the regions 4.5 ppm and 3.74 - 3.43 ppm respectively and these regions are obscured by the buffer water and DMSO in the samples.

As with Inhibitor 4, NMR spectroscopic methods were employed to investigate the interaction between NCS1 and Inhibitor 5. As previously mentioned the similarity between the two compounds only varies with the additional *para*-chloro group present on the phenyl ring. It was, therefore, surprising that with this small difference, Inhibitor 5 appeared to display what could be interpreted as a much stronger interaction with NCS1. This observation could be explained by the fact that the inclusion of chlorine makes the compounds more hydrophobic, which may fit in with overall characteristics of PPI binding (PPI interactions tend to be large, flat and hydrophobic in nature generally).

This conclusion is implied through a number of experimental results, firstly the acquisition of a 2-D ^{1}H ^{15}N HSQC spectra at 500 μM Inhibitor 5 and 50 μM NCS1 (Figure 3.17). The subsequent "chemical shift perturbation" calculations elucidated that the change in chemical shifts ranged from 0.00138 - 0.08 which is substantially more than the changes incurred through the effects of Inhibitor 4 at the same concentration.

Secondly there were a greater number of residues that were attenuated at 500 μM Inhibitor 5, 15 in total in comparison to the five induced through the effects of Inhibitor 4. Those 15 residues include; Asp37, Gln42, Tyr52, Lys53, Gly59, Arg94, Ala104, Asp126, Ile128I, Ile179I, Gln181, Leu183, Ser184 and Gly188, the position of the attenuated residues is not limited to a specific EF-hand region, although there are a large number found within the hydrophobic binding groove (Figure 3.20 a).

The top binding pose predicted through GOLD docking analysis for Inhibitor 5 is more linear and lies along the hydrophobic binding groove (Figure 3.20 b), in comparison to that of Inhibitor 4 which appears to adopt a curved pose (Figure 3.13 b). In light of this, through analysis of the attenuations or changes in chemical shift at 500 μM, three of those residues involved are of the original 6 that were used in the initial design process for the first generation of inhibitors (Chapter 2.1, Figure 2.2).

No residues found on the first α helix appear to undergo significant changes in chemical shift, an observation also seen with the effects of Inhibitor 4. There are a greater number of residues within the top 10 percentile located on EF-1 and EF-2, whereas those within the top 20 percentile tend to be located around EF-3 and EF-4. As the concentration of the inhibitor is increased to 1 mM (Figure 3.18 and 3.19), the range over which the changes in chemical shift of the residues occur, increases ever so slightly to between 0.006 and 0.105. Again as with Inhibitor 4, this is indicative of the fact that the protein has not yet reached saturation of inhibitor, despite the protein

to ligand ratio now being increased to 1:20. However as the difference between the two ranges is smaller than that of the difference between the two of Inhibitor 4, it could be inferred that the protein is reaching saturation faster with Inhibitor 5 than 4. Furthermore this could possibly indicate a stronger binding affinity, although this is not confirmed and could be done so through ligand titration analysis in future work.

Increasing the concentraton of Inhibitor 5 lead to the attenuation of a number of the residues that underwent a large change in chemical shift at the lower concentration of 500 μM, including Val68, Leu101 and Val180. It also increased a number of the "chemical shift perturbation" values for residues such as Asp44, Phe56 and Ser178. It did not however, affect the residues the regions that appeared to undergo the greatest number of significant changes EF-1 and EF-2. In addition, the non-specificity of interaction did not increase as the concentration of the inhibitor was increased (as seen with Inhibitor 4).

Following the first generation and Inhibitor 4, the effects of Inhibitor 5 on the aliphatic methyl and aromatic sidechain residues of NCS1 were also investigated. To do this as with the previous analysis, the difference between the control ^{1}H ^{13}C HSQC TOCSY spectra of 50 μM ^{1}H ^{15}N ^{13}C NCS1 and 0.5% DMSO with the ^{1}H ^{13}C HSQC TOCSY spectra of 50 μM ^{1}H ^{15}N ^{13}C NCS1 and Inhibitor 5 500 μM (Figure 3.21) are evaluated.

Inhibitor 5 appears to effect nine aliphatic methyl sidechain residues, of which only one residue is completely attenuated; Leu189 located on the terminal loop of EF-4. It does however, cause partial attenuation of eight residues, meaning that one of a pair of methyl groups are effected and those residues include; Leu_a16, Leu_b43, Val_a71, Ile_a128, Val_a132, Ile_a152, Val_a180 and Val_b190. Along with the attenuation of the residues, there are a number of notable changes in chemical shift including Ile51, Ile80, Ile124 and Ile179. As with previous inhibitors these residues are not located in a specific region of NCS1 but are in fact found across the four EF hand domains, on α helices and flexible loop regions. As with the first generation ligand Inhibitor 3 and Inhibitor 4, the aromatic Tyrε residues are un-attenuated by Inhibitor 5 (Figure 3.22), however residues that are attenuated include; Phe22, Phe67, Phe85, Tyr129, Phe153 and Phe169 primarily located in α helical regions of EF-2 and EF-4.

The effects of Inhibitor 5 on the aliphatic methyl and aromatic sidechain residues of NCS1, is in concordance with the 2-D ^{1}H ^{15}N HSQC spectra at the same concentration. We observe more changes in the NCS1 backbone and sidechain residues in the presence of Inhibitor 5, in comparison to the same spectra in the

presence of Inhibitor 4. The effects of both second generation inhibitors appear to be less than observed for Inhibitor 1 and Inhibitor 2, where we observed a large number of attenuations in sidechain residues.

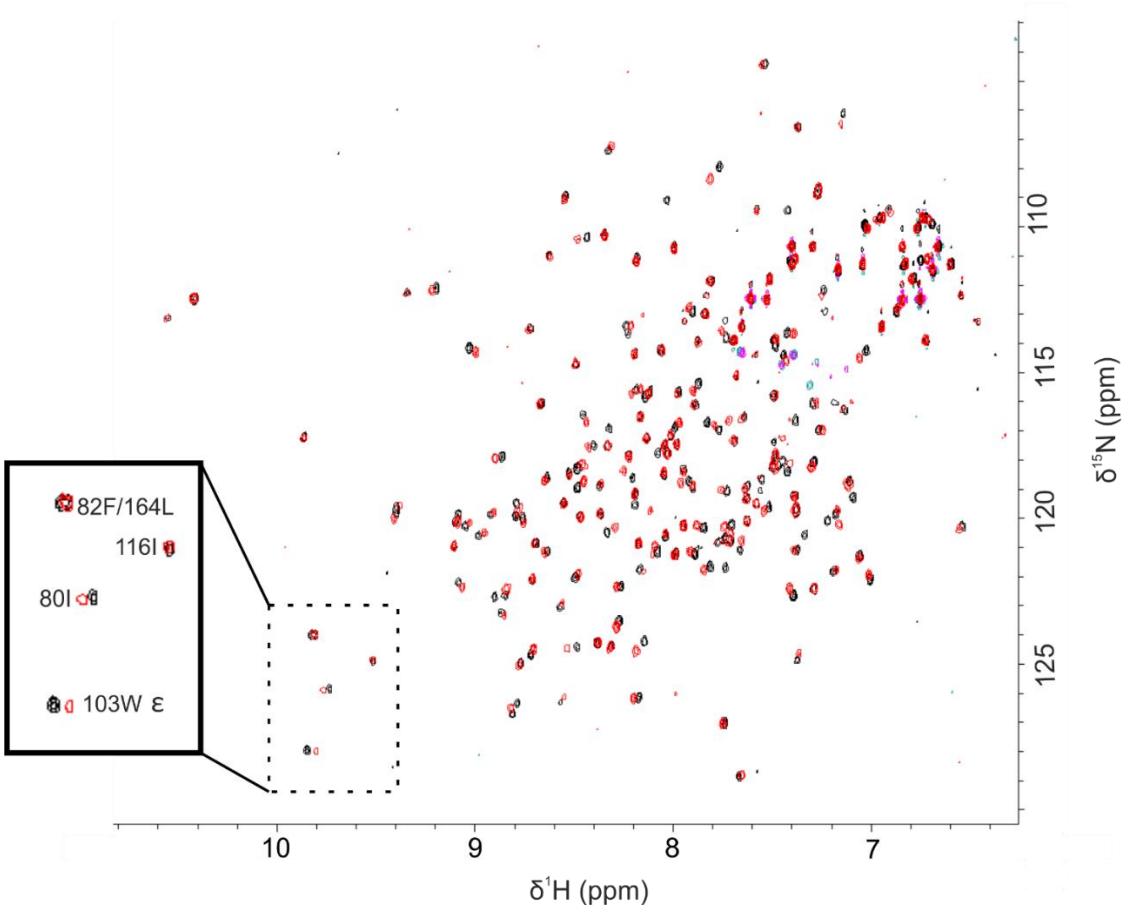


Figure 3.17- Ligand binding screen NCS1 with Inhibitor 5. A 2-D ^{1}H ^{15}N HSQC spectra of ^{1}H ^{15}N NCS1 50 µM and 0.5% DMSO (black), overlaid with ^{1}H ^{15}N NCS1 50 µM and Inhibitor 5 500 µM (red) P:L ratio of 1:10, ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 800 MHz spectrometer at 298K. Changes in chemical shift and attenuation of a number of peaks relating to a number of residues including Trp103 and Ile80 is indicative of an interaction between the inhibitor and the protein that appears to be stronger than observed for Inhibitor 4. The residues highlighted within the box were selected to indicate changes in specific residues within a conserved region of the HSQC spectra.

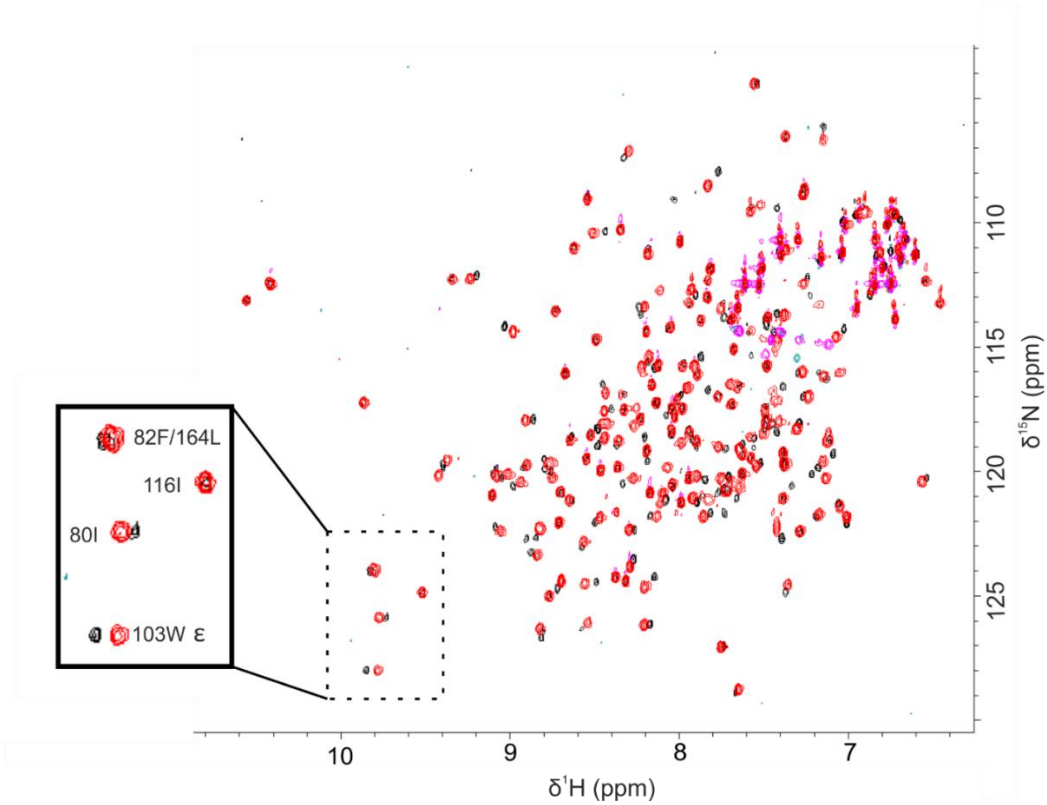


Figure 3.18- **Ligand binding screen NCS1 with Inhibitor 5.** A 2-D ^{1}H ^{15}N HSQC spectra of ^{1}H ^{15}N NCS1 50 µM and 1.0% DMSO (black) overlaid with ^{1}H ^{15}N NCS1 50 µM and Inhibitor 5 1000 µM (red) P:L ratio of 1:20, ^{1}H ^{15}N NCS1 50 µM prepared in NMR buffer pH 6.4, acquired on Bruker AVANCE II+ 800 MHz spectrometer at 298K. Further changes in the chemical shift of some of peaks Trp103 and Ile80 is further indication of an interaction between the inhibitor and the protein. It also highlights that increasing the concentration of the inhibitor is inducing further changes which is associated with the protein not being fully saturated at 500 µM. The residues highlighted within the box were selected to indicate changes in specific residues within a conserved region of the HSQC spectra.

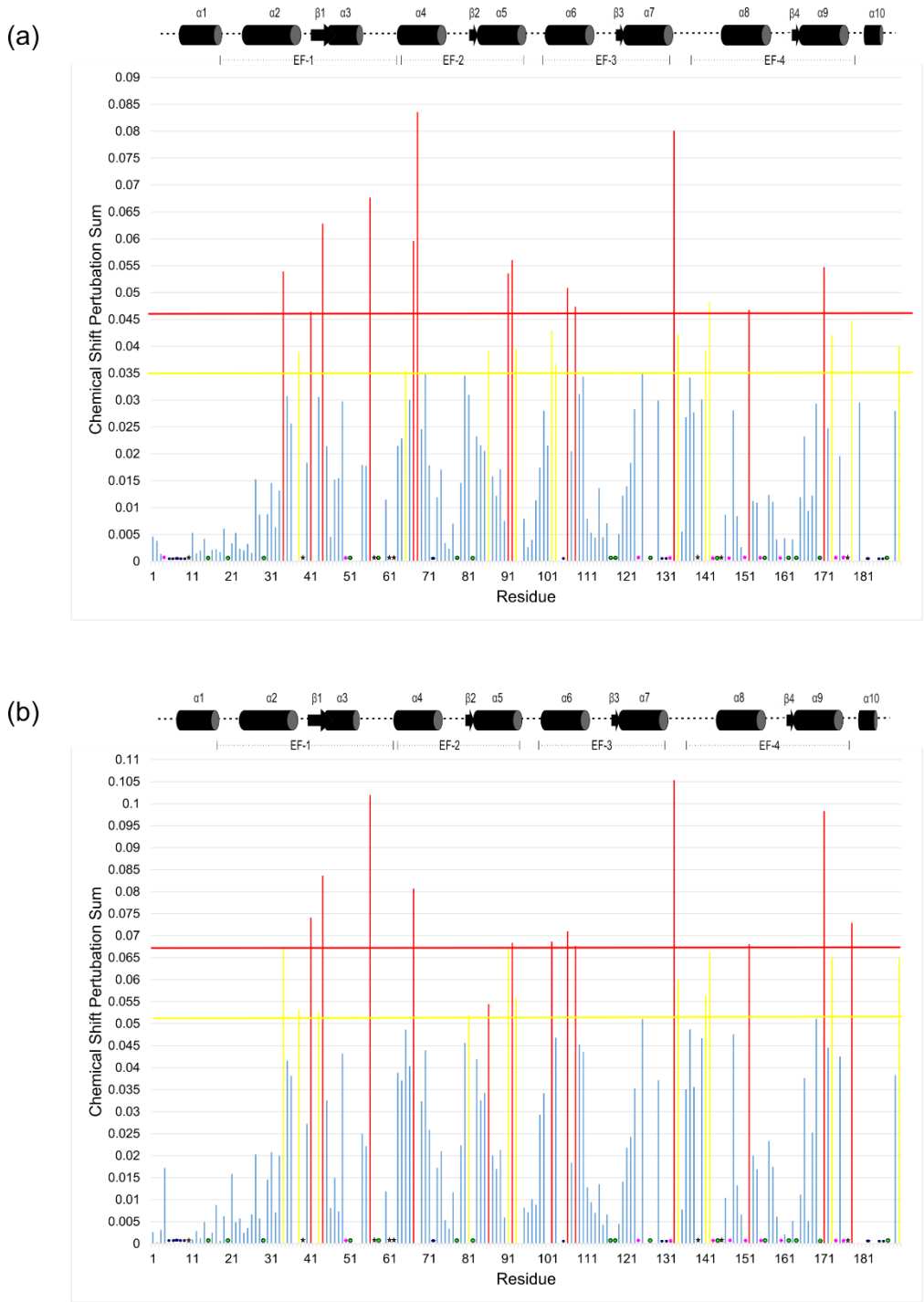


Figure 3.19– (a) Histogram of the perturbations of NCS1 residues from ligand binding screen NCS1: Inhibitor 5 at 500 μM, P:L 1:10. The changes in chemical shift pertubation for each residue was subsequently calculated, and the difference changed into percentile values of the total number of residues affected for analysis. Those residues whos chemical shift underwent a change in ppm within the top 10 percentile can be seen above the red line and those residues include: Val68, Gly133, Phe56, Asp44, Phe67, Thr92, Glu171, Phe34, Val91, Lys106, Glu142, Tyr108, Ile152. The yellow line represents those residues who's change in chemical shift was within the top 20 pecentile and includes Gly41, Ser178, Arg102, Asn134, Ser173, Val190, Ser93, Glu141, Ile86, Cys36, Trp103, Ala65, Val125. **(b)** Histogram of the perturbations in NCS1 residues from ligand binding screen NCS1: Inhibitor 5 at 1 mM P:L 1:20, those residues whos chemical shift underwent a change in ppm within the top 10 percentile can be seen above the red line and those residues include: Gly133, Phe56, Glu171, Asp44, Phe67, Gly41, Ser178, Lys106, Arg102, Thr92, Ile152, Tyr108, Phe34. The yellow line represents those residues who's change in chemical shift was within the top 20 percentile and includes: Val91, Glu142, Val190, Ser173, Asn134, Glu141, Ser93, Ile86, Cys38, Leu43, Glu81, Phe169, Val125. ● Unassigned residues ✿ Single peaks assigned with two residues ★ Proline ✸ peaks that became too broad to assign.

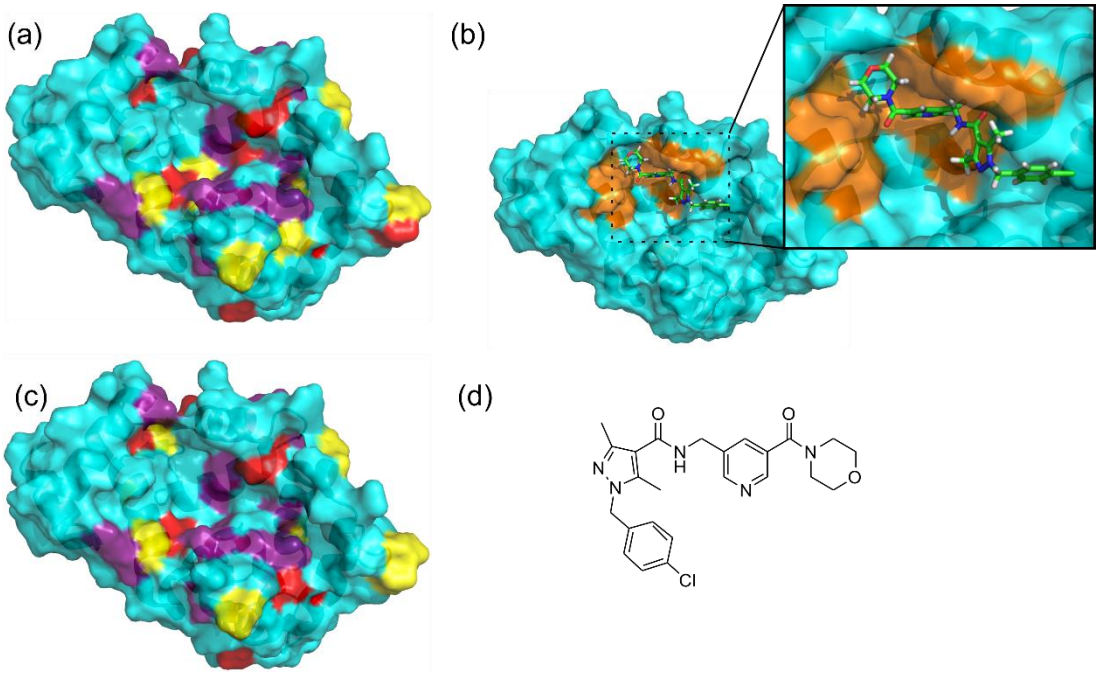


Figure 3.20– (a) Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 5 at 500 μM determined from the corresponding histogram the residues within the top 10 percentile coloured in red and top 20 percentile in yellow, those coloured in purple belong to the residues that were attenuated upon the addition of the ligand and not due to the affects of DMSO (PDB 5AER). **(b)** Prediction of the binding pose for the interaction between Inhibitor 5 and NCS1 (3.1.2), those key residues highlighted in orange.[4] **(c)** Crystal structure of NCS1 in cyan with the corresponding residues perturbed by interactions with Inhibitor 5 at 1 mM determined from the corresponding histogram, the residues within the top 10 percentile coloured in red and top 20 percentile in yellow, those coloured in purple belong to the residues that were attenuated upon the addition of the ligand and not due to the affects of DMSO (PDB 5AER). **(d)** Chemical structure of Inhibitor 5.

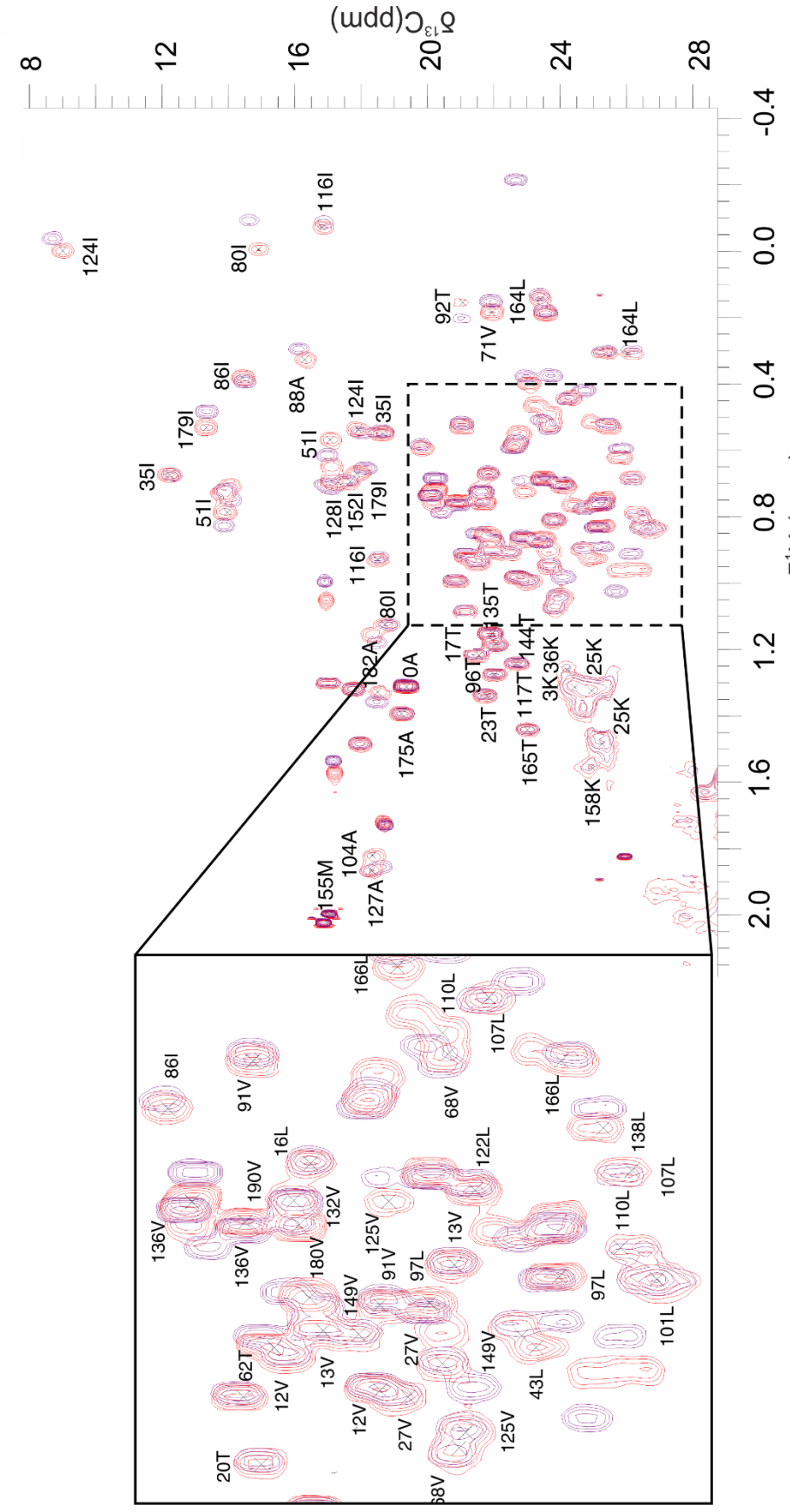


Figure 3.21- An overlay of two methyl sidechain regions of NCS1 with and without Inhibitor 5. The ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 0.5% DMSO (purple), overlaid with the ^{1}H ^{13}C HSQC spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 5 500 µM (red). Those residues of NCS1 related to the Inhibitor 5 spectra for which the transfer of assignment was possible can be seen, there a number of notable changes in chemical shift including Ile51, Ile80, Ile124 and Ile179 as well as the attenuation of residues such as Leu89. Both spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

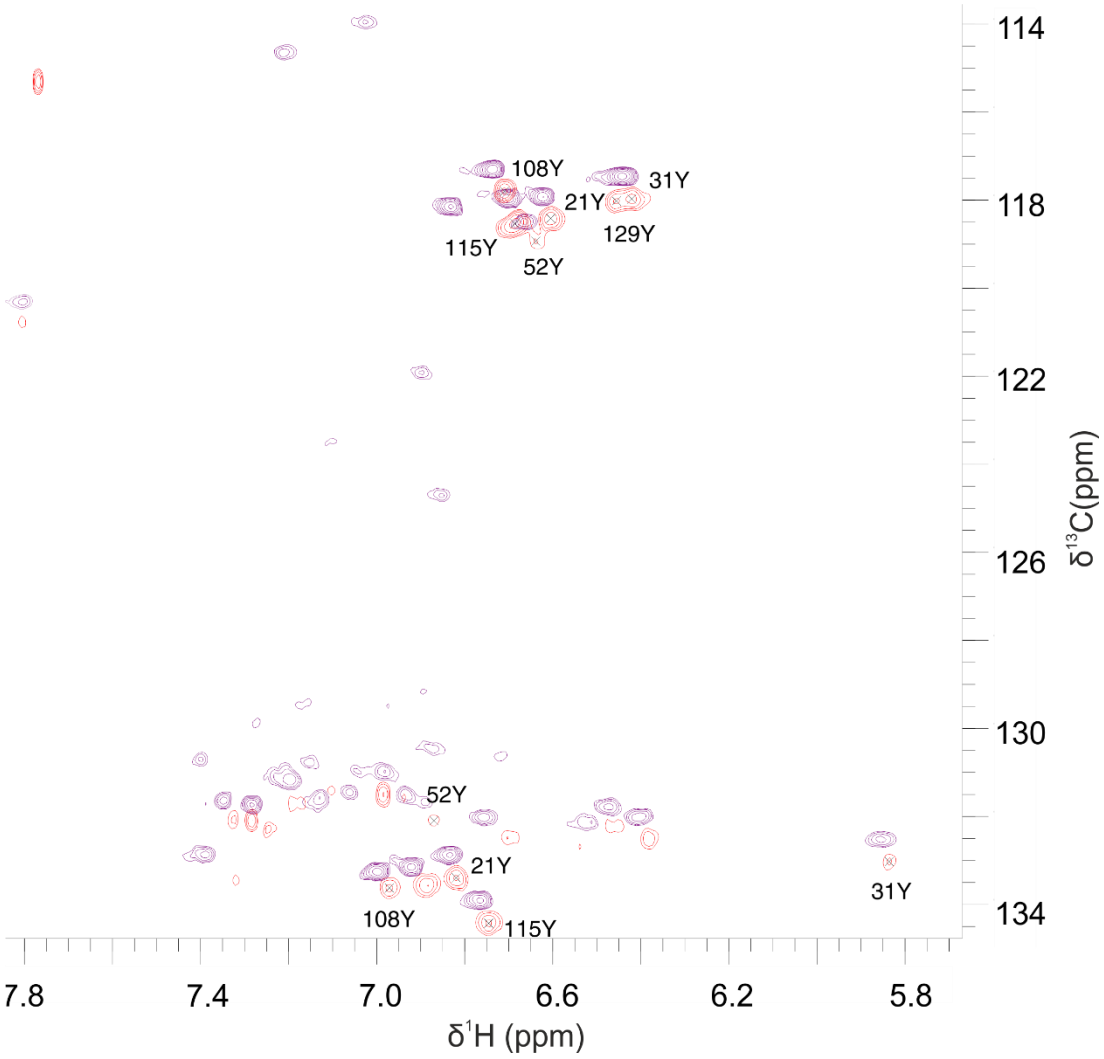


Figure 3.22- Two overlaid aromatic sidechain spectra of NCS1 with and without Inhibitor 5. The aromatic ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and 0.5% DMSO (purple), overlaid with the ^{1}H ^{13}C spectra of ^{1}H ^{15}N ^{13}C NCS1 50 µM and Inhibitor 5 500 µM (red), the aromatic residues labelled are those for which the assignment was transferrable. Those residues for which transfer of assignment was not possible include those which reside on the same peak, or those which were attenuated by the presence of Inhibitor 5 including; Phe22, Phe67, Phe85, Tyr129, Phe153 and Phe169. The spectra were acquired at 298K in NMR buffer pH 6.4 on Bruker AVANCE II 600 MHz spectrometer.

As discussed previously the use of the 2-D ^{1}H ^{15}N HSQC experiment allowed the interpretation of the specific amino acids within NCS1 that underwent a change in chemical shift in the presence of Inhibitor 5. However, other NMR experiments such as the STD and waterLOGSY experiments enable us to investigate what parts of the inhibitor interacting with the protein. By combining the two sets of data a more detailed understanding of the interaction between the protein and the ligand can be determined.

As with Inhibitor 4 the 1-D ^{1}H saturation transfer resonance (STD) and water ligand observation with gradient spectroscopy (waterLOGSY) data for Inhibitor 5 provided quite conclusive indications of a binding event, and identified the regions of the inhibitor which make contact with the protein. As explained with Inhibitor 4, there is

no ligand signal in the STD spectra of Inhibitor 5 on its own in NMR buffer, as there is no protein from which the magnetisation can be transferred.

Upon closer inspection of the spectra (Figure 3.23) and comparison to that of the protein containing sample the ligand signals relating to resonances 1, 2, 3, 4, 5, 8, 9 and 10 are strong and visible. Resonances 7, 11 and 12 are not observed as they appear in the regions 4.5 ppm and 3.74 - 3.43 ppm respectively and these regions are obscured by the buffer, water and DMSO in the samples. Examination of the results of the waterLOGSY spectra in parallel to that of the STD spectra, the resonances relating to positions 4 and 5 are inverted, and the resonances relating to positions 1, 2, 3, 8, 9 and 10 are significantly attenuated. Both spectra strongly indicate that there is a binding event occurring between Inhibitor 5 and NCS1 and that the majority of the protons of the inhibitor are contributing to this binding interaction.

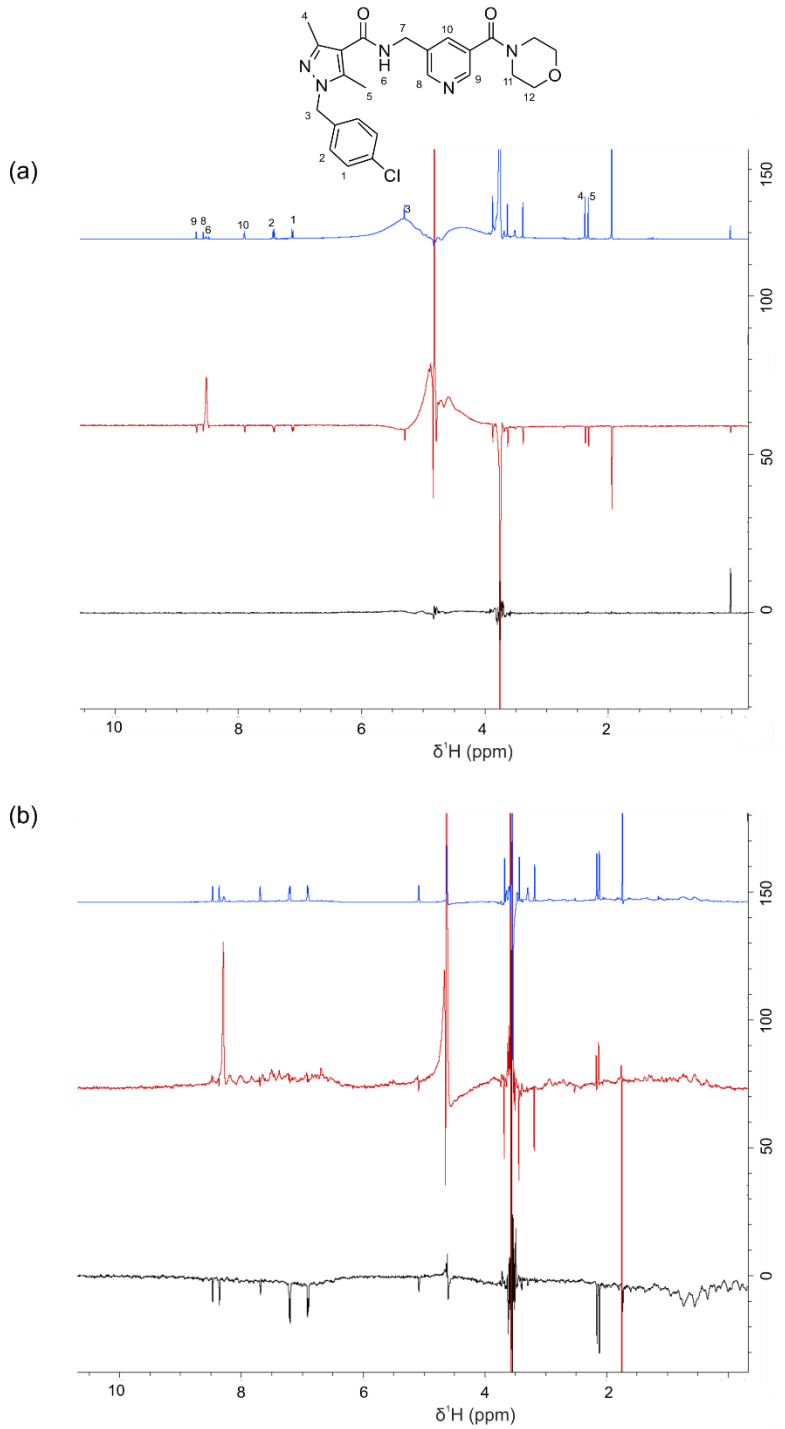


Figure 3.23- 1-D ligand screening spectra of Inhibitor 5. (a) Buffer standard 500 µM Inhibitor 5 in NMR buffer; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(b)** Binding of 500 µM Inhibitor 5 to NCS1 50 µM in NMR buffer; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). Comparison between the buffer controls (a) and the protein containing experiments (b) indicate possible binding events occurring at ligand positions 1, 2, 4, 5, 6, 8, 9, 10. This is deduced in the STD spectra where the relative ligand signals appear strong and negative, the corresponding peaks in the waterLOGSY spectra for positions 1, 2, 8, 9 and 10 are completely reduced in intensity with peaks 4 and 5 flipped on the opposite phase compared with the other resonances. The results of both experiments are indicative of a relatively strong binding interaction. Resonances 7, 11 and 12 are not observed as they appear in the regions 4.5 ppm and 3.74 - 3.43 ppm respectively and these regions are obscured by the buffer water and DMSO in the samples.

3.3.2 Binding affinity determination by ITC

General experimental procedures

The experimental procedure and the conditions used for the second generation characterisation using ITC experiments conducted in this section are discussed in detail in Chapter 6.3.2.9. Briefly, all unlabelled NCS1 protein samples were prepared in the ITC buffer consisting of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ at pH 7.5. The concentration of the protein and ligand were kept constant for the initial experiments and subsequently varied as required.

Due to the improved solubility of the inhibitors the experiments were conducted in the classical manner meaning that the ligand was titrated into the cell containing the protein NCS1. As discussed previously, experimentally ITC allows the determination of the thermodynamic properties ΔH and ΔS for a paticular binding interaction, where ΔH is the enthalpy change and ΔS the change in entropy. These properties can be interpreted, providing information on the changes occuring between two binding partners. The details behind how the thermodynamic properties can be used to derive further information of the binding interaction has been discussed in detail previously in Chapter 2 and can be found in Section 2.3.5. Analysis of the raw data was conducted in the software package Origin7 (OriginLab, Northampton, MA) and involved subtraction of a respective DMSO control experiment from the experimental data of each ligand.

Due to the weaker nature of the interaction between Inhibitor 4 and NCS1 demonstrated using NMR spectroscopy, only the binding affinity of the interactions between Inhibitor 5 and NCS1 were investigated using ITC. The experimental results can be found in this section.

3.3.2.1 Inhibitor 5

With the improved solubility of the second generation of inhibitors, the experimental techniques used were able to follow more classical approach, the inhibitor titrated into the cell containing the protein.

In the case of Inhibitor 5, a number of experiments were performed at different concentration ratios between the Inhibitor and protein, an optimised isotherm can be seen (Figure 3.26 a). In this experiment Inhibitor 5 at a final concentration of 8 mM was titrated into NCS1 400 μM, the experiment was conducted at two temperatures 25 ˚C and 10 ˚C. Interestingly, variation of the temperature resulted in two different binding constants, at 25 ˚C; the K_D was four times smaller than that at 10 ˚C, despite the thermodynamic data being similar (Table 3.2, Figures 3.24, 3.25).

Table 3.2- Inhibitor 5 ITC data. A table showing the thermodynamic properties from the ITC experiment of 8 mM Inhibitor 5 and 400 µM NCS1 at two different temperatures. Both sets of thermodynamic properties are similar although the derived K_D is four fold higher at the higher temperature with both experiments having an improved goodness of fit in comparison to the first generation of inhibitors.

Experiment Title	ΔH (kcal mol^{-1})	-TΔS (kcal mol-1)	ΔG (kcal mol-1)	K_D µM	Chi^2
10˚C Inhibitor 5 8mM: NCS1 400µM	-0.738 (±0.03912)	-4.87	-5.61	47.2	2262
25˚C Inhibitor 5 8mM: NCS1 400µM	-1.012 (±0.135)	-4.02	-4.82	201.207	400.8

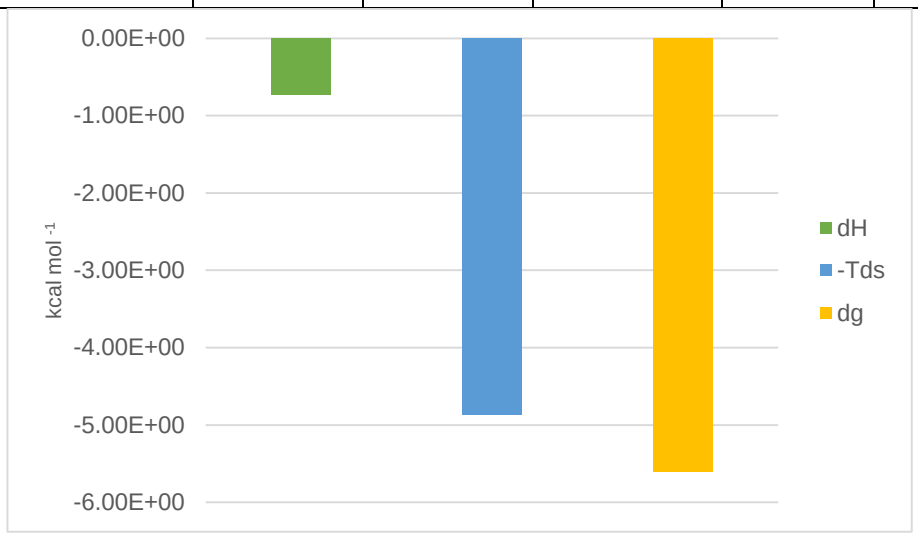


Figure 3.24- The thermodynamic parameters (dH= ΔH green, -TdS= -TΔS blue and dG= ΔG yellow) obtained from the ITC experiment of the binding interaction between Inhibitor 5 8 mM and NCS1 400 µM at 10 ˚C. When determining ΔG the most influential value is ΔH, ideally it should be large and negative, -TΔS contributes substantially less to the value of ΔG and the binding affinity and tends to be small and positive or small and negative.

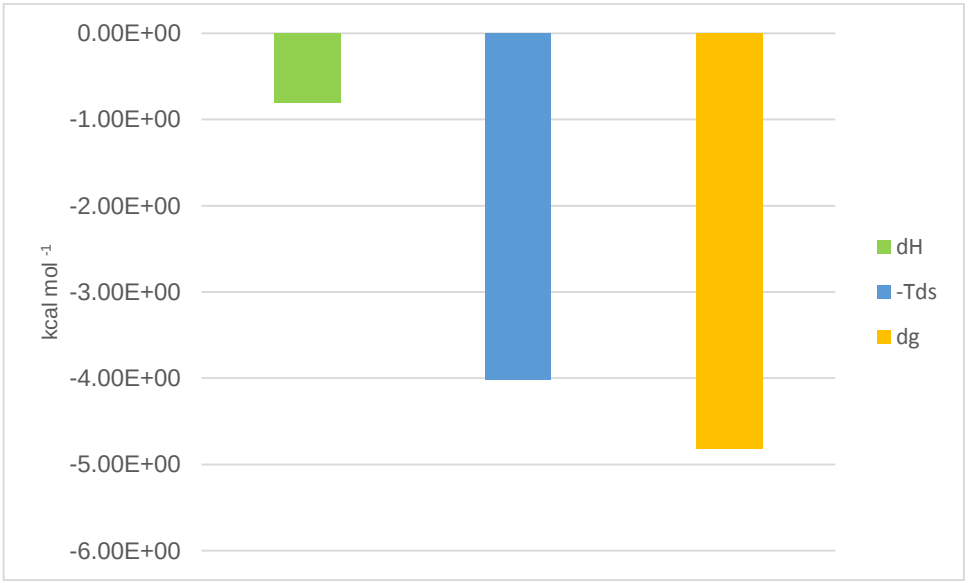


Figure 3.25- The thermodynamic parameters (dH= ΔH green, -TdS= -TΔS blue and dG= ΔG yellow) obtained from the ITC experiment of the binding interaction between Inhibitor 5, 8 mM and NCS1 400 µM at 25 ˚C.

When we compare the two isotherms more closely, we can see the goodness of fit value (Chi^2) for the experiment at 25 ˚C is more favourable as it is smaller.[212–214] Both experiments exhibit a negative T∆S value which can be inferred as hydrophobic interactions occurring, this data agrees with the nature of the NCS1 binding site and indicates the reaction is entropically driven. There is also a difference in the stoichiometry of binding (n) at 10 ˚C the ligand: protein ratio is 2:1, however increasing the temperature results in this stoichiometric ratio changing to 1:1. The most likely explanation is that at the lower temperature NCS1 is a dimer, hence presenting two binding sites for Inhibitor 5.

At both 10 ˚C and 25 ˚C neither experiments appear to reach saturation, infact continuation of each experiment followed by concatenation of the relative data sets revealed a major problem with using ITC to determine the binding affinity for this system. The delay in injection between the first experiment and starting the second meant that the residual DMSO already in the cell had the chance to interact with NCS1. This therefore resulted in an abnormal change in heat given off when the experiment was continued (concatenated figures of the experiment at 10 ˚C can be found in the appendix, Section A.4).

Therefore, although the solubility of the inhibitors has improved allowing for a higher concentration to be used within the experiments, the percentage of DMSO used is also increased to a much greater percentage in comparison to that used for first generation characterisation. The interactions of DMSO with NCS1 at this higher concentration, mean that ITC is an unsuitable technique for determining the binding affinity of the system, until the requirement for DMSO is diminished further.

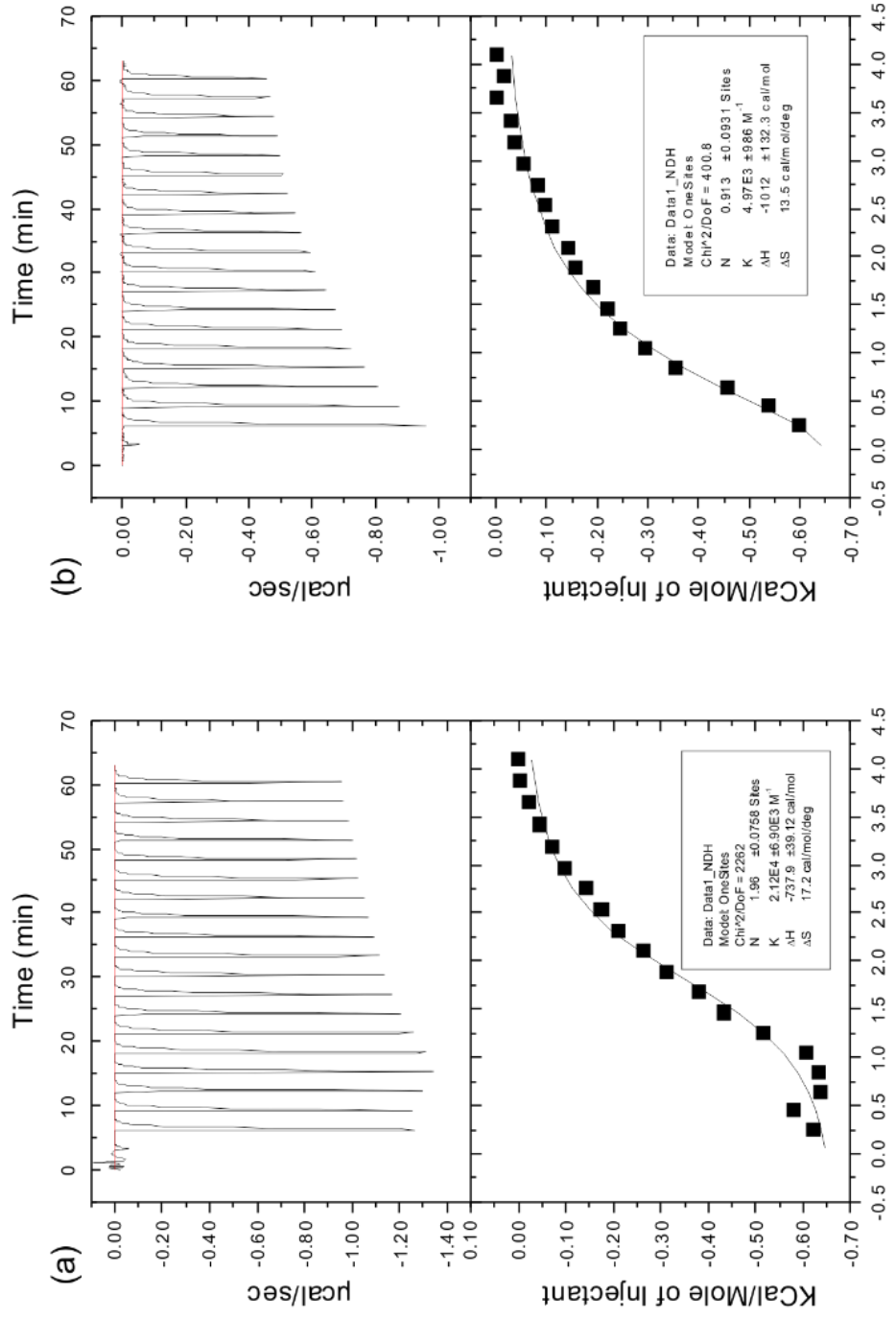


Figure 3.26- Binding affinity determination, Isothermal Titration Calorimetry curves of NCS1 and Inhibitor 5. Two binding isotherms, ITC Isotherm (top) and resultant curve (bottom) from the experiment 8 mM Inhibitor 5 and 400 µM NCS1 at 10 ˚C **(a)** and 25 ˚C **(b)** 8% DMSO, 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5. Fitting of the curve was conducted in Origin 7 using the single set of sites curve fitting model produced a binding model with a derived binding affinity K_D of 47.2 µM and 201.2 µM for the experiments at 10 ˚C and 25 ˚C respectively. Both experiments have improved Chi^2 values relating to the goodness of fit.

3.3.3 Binding affinity determination using Fluorescence Spectroscopy

General experimental procedures

All fluorescence data of the second generation of inhibitors was collected at The University of Liverpool on a Cary Eclipse Fluorescence spectrophotometer (Agilent technologies), the raw data was analysed in Origin 7 (OriginLab, Northampton, MA).

Unlabelled NCS1 was prepared as described in Chapter 6.3.2. Briefly, unlabelled NCS1 was buffer exchanged into fluorescence buffer consisting of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5; and aliquoted into 11 samples at a final concentration of 1 µM (v/v) (200 µL).

For the individual single point reads of the tryptophan fluorescence assay, Inhibitor 4 and Inhibitor 5 were added in the required volume from a single stock solution in 100% DMSO, to the already aliquoted NCS1 samples contained in eppendorf's. The concentration range for each inhibitor over the 11 samples was 0 - 1 mM. Each individual assay sample was transferred to a quartz 16.160-F/Q/10 cuvette (Starna Scientific) and excited at a wavelength (λ) of 280 nm, with the excitation slit width 5 nm and the emission slit width 20 nm.

3.3.3.1 Inhibitor binding assay

In keeping with previous fluorescence assays performed on the first generation of inhibitors, the interactions between inhibitors 4 and 5 with NCS1 were investigated initially in an identical manner. However it was apparent that the concentration range for the inhibitor of 0 - 90 µM should be increased as saturation was not achieved in this range, therefore the maximal concentration of inhibitors 4 and 5 were increased to 1 mM. At this maximal molar excess of inhibitor, the protein ligand ratio (P:L) was 1: 1000.

As previously explained, intrinsic tryptophan fluorescence is highly sensitive to any changes in the local environment. It is therefore a useful technique for evaluating changes in a proteins structure, induced through possible binding interactions and also in determining binding affinity.[216,294] The pre-requisite for the experiment is the presence of tryptophan residues, especially if like in NCS1 they are located in any key binding domains. In contrast to the first generation of inhibitors, it is apparent that both Inhibitor 4 and Inhibitor 5 had less of an effect with the fluorescence of NCS1. This can be interpreted from the fact that the inhibitors were unable to cause saturation of the tryptophan fluorescence of NCS1 at 90 µM (a 1:90 P:L ratio) and even when the concentration was increased to 1mM, Inhibitor 4 was still unable to cause saturation despite being in such great excess. In fact Inhibitor 4 was unable to

cause a significant decrease in the fluorescence until it was at a concentration of 62.5 μM (P:L 1: 62.5).

The increasing concentrations of Inhibitor 4 and 5 also did not appear to cause any structural changes of the protein (Figure 3.27) as observed in the experiments of inhibitors 2 and 3 (Chapter 2.3.8, Figure 2.49). However, these results do appear to be in concordance with the NMR data which suggests that Inhibitor 4 does not cause a significant change in the chemical shift of either Trp30 or Trp103 at a 20 fold molar excess.

Inhibitor 5 does cause a more significant change in the chemical shift of Trp103 although it is not attenuated, when comparing this observation to the tryptophan fluorescence we can infer that although saturation is not reached, it does have a greater effect on the tryptophan fluorescence of NCS1 than Inhibitor 4. This does not mean that the inhibitors are not interacting with the protein, it simply means that any interactions may not involve the tryptophan residues.

Binding affinity of the inhibitors using tryptophan fluorescence was not determined because as with the ITC data, the use of DMSO to solubilise inhibitors 4 and 5 interfered with the fluorescence readings. Any control experiments found that this interference changed over time and hence could not be used to normalise the data, therefore any binding constants derived from the experiments would be unreliable.

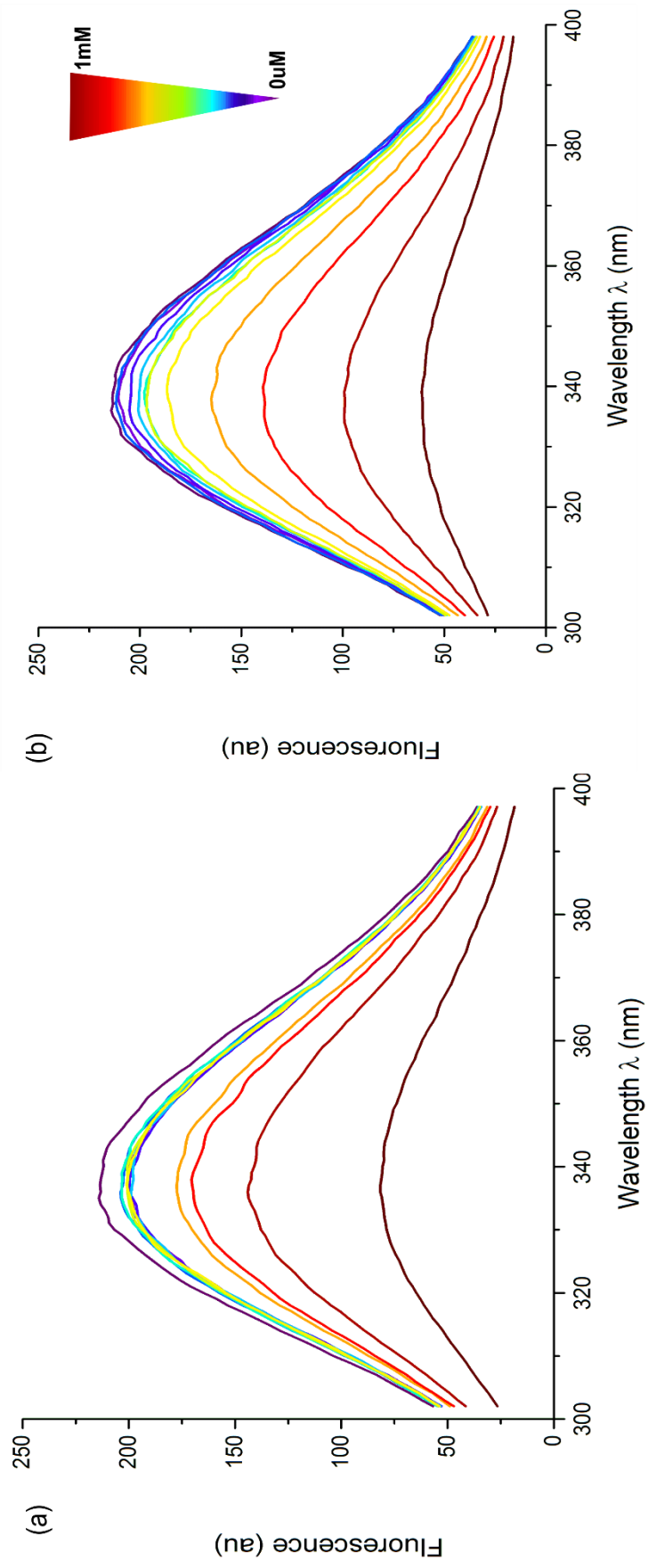


Figure 3.27- Fluorescence Binding titration assays of NCS1 1µM with Inhibitors4 and 5 over a range of 0 - 1 mM. (a) Inhibitor 4 fluorescence binding assay, NCS1 at 1 µM and Inhibitor 4 titrated in to the individual protein samples at concentrations 0 µM, 0.9 µM, 1.9 µM, 3.9 µM, 6.7 µM, 15.6 µM, 31.25 µM, 62.5 µM, 125 µM, 500 µM and 1 mM. **(b)** Inhibitor 5 fluorescence binding assay, NCS1 at 1µM and Inhibitor 5 titrated in to the individual protein samples at concentrations as with Inhibitor 4. In both cases we observe a quenching in the fluorescence.

Summary

The structure-based development of inhibitors to target the protein-protein interaction between NCS1 and D2R peptide has involved the biophysical analysis of a first generation of molecules, the results of which subsequently led to the development and synthesis of a second generation.

Designed around the most successful inhibitor of the first generation (Inhibitor 2) inhibitors 4 and 5 displayed an improved solubility profile (Figure 3.7), which aided the biophysical analysis of their interactions with NCS1. The analysis of the interactions of the inhibitors with NCS1 using NMR spectroscopic methods, indicated that the additional *para*-chloro group of Inhibitor 5 appeared to improve the interaction. This can be inferred from comparisons of the 2-D ^{1}H ^{15}N HSQC spectra (Figures 3.10, 3.11, 3.17, 3.18) where we observe, a greater number of changes in chemical shift of NCS1 residues in the presence of Inhibitor 5, than at the same concentrations of Inhibitor 4.

Moreover when we subsequently compare this data to that of the first generation (Chapter 2.3.4, Figures 2.28, 2.29, 2.34, 2.39), we can deduce that inhibitors 4 and 5 display an improved interaction with NCS1. Furthermore there appears to be a greater number of residues that undergo a significant change in chemical shift around the hydrophobic binding groove (Figures 3.13 and 3.20), than seen with Inhibitor 3 or even the parent compound Inhibitor 2 (Chapter 2.3.4, Figure 2.34).

Despite the improved solubility profiles of Inhibitors 4 and 5, determination of the binding affinity using ITC techniques was still not reliable. It was ascertained that solubilising the compounds in DMSO caused difficulties with the results. This is due to the fact that to reach the necessary concentration of Inhibitor 5 required to saturate NCS1 binding, the large amount of DMSO (8%) had a detrimental impact on the experimental readout. DSMO was found to interact with the protein in a time dependent fashion, altering the thermodynamic results, such that any derived binding constant (K_D) was not a clear measure of binding. A similar phenomenon was observed in the tryptophan fluorescence experiments. Future work to achieve an accurate binding constant using ITC and fluorescence techniques, could include using a different organic solvent that does not interact with NCS1 to solubilise the inhibitor further increase solubility so DMSO is not needed.

Chapter 4
Small Molecule Compound Library

.

4. Small Molecule Compound Library

4.1 Introduction- The role of NMR and computation in fragment-based drug discovery.

There are a number of different biochemical techniques often used in studying protein-ligand interactions including isothermal titration calorimetry (ITC), biosensor array systems (surface plasmon resonance (SPR) and Octet system), microscale thermophoresis, fluorescence polarisation (FP), fluorescence resonance energy transfer (FRET), enzyme linked immunosorbent assay (ELISA), nuclear magnetic resonance (NMR) and X-ray crystallography. Each of these methods has its strengths and weaknesses and these are summarised in Table 4.1.

The use of NMR spectroscopy to probe protein-ligand interactions is an important field in drug discovery and development, often used by research groups in fragment-based drug discovery processes (Figure 4.1).[295] In some cases libraries containing many thousands of small molecules are screened in a high throughput manner to discover novel hit compounds, the process can be both costly and time consuming, involving numerous multifaceted steps.[295] To avoid wasting time and resources it is important for the design of the compound libraries to be well thought out, taking into account the likelihood of fragment: target interaction, diversity of the library ADMET properties such as solubility to reduce the number of false positives and enhance developability potential of any hits, before the experimental screening is undertaken.[296]

The advances in computing mean that the time it takes to perform computational screening has become significantly reduced compared to a decade ago, making it a much more efficient process. However the reliability with discovering hit compounds using this method in comparison to high throughput library screening is not as successful.[297] As with the advances in computing, the knowledge of the structures of target proteins is also increasing, to date the number of distinct protein sequences in the PDB stands at 35639 and of those the structures of human sequences 28941.[298] Thus to improve the probability of finding quality hit molecules, computational techniques are used in combination with the known target structures in virtual screening processes such as molecular docking.[297]

Table 4.1- A table comparing the advantages and disadvantages of a number of the biophysical techniques used to study protein- ligand interactions.

Technique	Description	Advantages	Disadvantages
ITC [210]	The direct measurement of the heat generated or absorbed during the interaction of two molecules.	Able to determine binding affinity. Label free. No molecular weight limitations. Optical clarity un-important.	Requires estimate of K_D to calculate concentrations of protein and ligand to use for initial experiment. Can only measure K_D values between mM - nM, weaker binding interactions require a specific "low c-value" titration. Sensitive to differences in buffer, resulting in artificial heats. Sensitive to bubbles in the protein and any precipitation of protein or ligand.
SPR [299]	The label-free detection of bio-molecular interactions using polarised light.	Label free. Good for evaluation of macromolecules. Equilibrium measurements ΔH and K_D. Able to detect low affinity interactions. Requires a small amount of protein.	Not good for high throughput screening (HTS) and concentration assays. Optimal for large molecules > 1000 Da. Lengthy assay development.
Octet System [300]	Optical analytical technique for the real time analysis of biomolecular interactions.	Optimised version of SPR technique able to withstand HTS whilst maintaining accuracy. Label free	
Microscale Thermophoresis [301]	The quantitative analysis of protein interactions in free solution. Using the directed motion of molecules in temperature gradients this is an all optical approach with fluorescence excitation.	Can use sub-nM concentrations of protein. No need to immobilise. Measurements can be selective for the specific fluorescent label. Not limited by the molecular weight ratio of binding partners. Sensitive to low affinity interactions.	Requires fluorescent labels.
Fluorescence Polarisation [217]	The use of plane polarised light to provide information on molecular orientation and mobility as well as the processes that modulate them such as protein-ligand interactions.	Fast, accurate quantitative measurements therefore good for HTS. No need for solid supports. Sensitive to pico-molar range. No hazardous radioactive waste generated. Real-time measurements for kinetics. Insensitive to variations in concentration.	Requires fluorescent labels.
F.R.E.T [302]	Detects the proximity of fluorescently labelled molecules over 10 - 100 Å.	Selective to those molecules within 10 - 100 Å of the protein.	Requires fluorescent labels. The absorption spectrum of the acceptor must overlap with that of the donor.

ELISA	A type of solid phase-enzyme assay to measure the presence of an analyte. Can be used to detect binding interactions between a protein and ligand.	Label free. Can be used for weak binding interactions. Good for HTS once optimisation complete.	Requires specific antibody for protein. Requires immobilisation. Requires optimisation of assay.
X-ray crystallography [303]	Used to identify the atomic and molecular structure of a crystallised protein: ligand complex.	A well-established technique. Data processing is highly automated. Can be used for large molecules.	Can only be used to determine the structure of a protein-ligand complex not the kinetics of the interaction. Protein has to be able to form stable crystals that can diffract. Optimisation of crystallisation conditions for each protein: ligand complex can be time consuming.
NMR	The phenomenon of Nuclear Magnetic Resonance is used to determine protein ligand interactions as well as the affinity of binding.	Reliable well established technique. Very sensitive. Able to determine K_D of binding and the structure of the protein: ligand complex. Can be used for HTS.	Requires higher concentrations of protein, which may lead to aggregation. Requires libelling such as ^{15}N and ^{13}C. Can be labour intensive.

There are many different docking programmes which operate a conformational and positional searching and scoring regime, the different poses of each ligand in the target protein can be proposed, scored and then ranked.[238] These binding predictions are then combined with calculations of how the compounds may behave *in vivo* (eg: logD) if they are to be developed for therapeutic use. However these computational simulations only aid the identification of compounds that make up a library, which must then subsequently undergo an actual screening process using a variety of biochemical and or biophysical techniques. Following computational screening, the next step is often experimental screening using one or more of the methods listed above (Table 4.1). Of these, the NMR method is widely used for fragment-based screening since this method is amenable for detection of binding over a very large affinity range; there are different NMR experiments which are suitable for the different affinities.

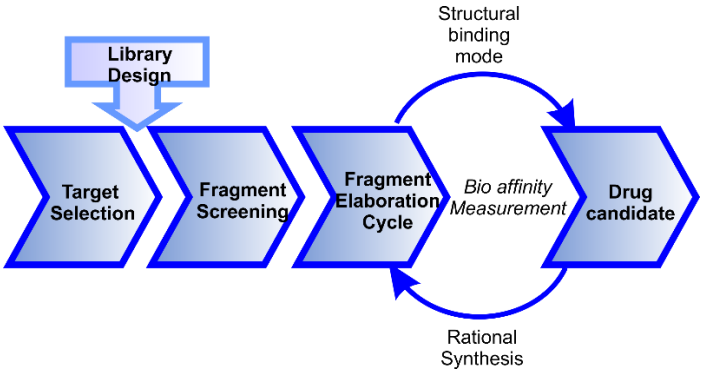


Figure 4.1– An example of a fragment-based drug discovery pipeline- The general workflow pipeline involved in a fragment-based approach. Target selection and library design enables the fragment screening process using a number of biophysical techniques, to determine hit compounds. Which are taken through an iterative optimisation process, involving determination of the binding mode along with affinity measurements resulting in possible drug candidates (adapted from [295]).

4.1.1 NMR method for ligand screening

The use of NMR in the screening process has developed over a number of years; some initial investigative studies used NMR for hit lead validation through identification of interactive sites.[304] The original protein target was FKBP, a protein known to bind to FK506, inhibit calcineurin and block T-cell activation.[304] The protein was known to have two interacting sites and so the fragment-based approach had two targets.[304] The SAR process used fragment screening and optimisation processes to screen and optimise an initial ligand to fit the first target site and then screened a second set of fragments to fit the second site.[304] These fragment compounds were then linked to create a larger compound with a K_D of 100 nM.[304] Due to the low molecular weight nature of fragment-based drug design, the biophysical techniques used to screen the fragment-based libraries need to be able to detect weak binding interactions.[295]

There are numerous experiments that can be used to do this such as SPR, thermal shift assays (TSA) and NMR experiments.[295] These include 1-D ^{1}H saturation transfer difference (STD) [305] and 1-D ^{1}H water ligand observation with gradient spectroscopy (waterLOGSY).[306]

The STD experiment was originally developed in 1979 for resonance assignments of a haem group within and enzyme; it was later used to characterise the interaction of a ligand binding to a receptor.[307,308] The experiment uses a sample containing protein with the ligand present in a very large excess (typically over 50:1 ligand: protein ratio). During the NMR experiment, the protein protons at a resonance position clear of ligand signals are irradiated with a radio frequency (upfield spectral regions above 0 ppm are normally used due to a lack of ligand signals in these regions). The magnetisation experienced by that single protein resonance then

diffuses throughout the protein; if the ligand is bound then intermolecular transfer of the magnetisation to the bound ligand occurs, which in turn transfers the magnetisation to the free ligand in solution (Figure 4.2). As a result this appears as strong ligand signals that are visible in the 1-D ^{1}H difference spectrum and it allows us to visualise which parts of the small molecule are binding to the protein. The relative intensity of ligand signal in an STD spectrum is not dependent on the strength of binding; rather, it is dependent on the "off rate" of the ligand under steady state equilibrium between the free and bound state of the ligand.[291] As such the STD method is suitable for weakly binding ligands because of their fast off rates. The off rate (K_{off}) is related to the affinity through the equation 4.1, where the on rate (K_{on}) is diffusion controlled around 10^{-9} $M^{-1}S^{-1}$. We can therefore say that K_{off} is directly related to the affinity and hence the STD experiment will give an indication as to if there is weak or strong binding.

$$KD = \frac{Koff}{Kon} \qquad 4.1$$

A variant of the STD experiment is the waterLOGSY, it is a technique that is often used in place of the STD experiment, in those situations where the protein may be prone to aggregation. The experiment relies on the bulk water which is the water found to bind at protein ligand interfaces and is often known to surround ligands.[292,293] With this experiment the "on resonance" saturation is applied to the frequency at which water is found to occur, as with the STD if the ligand is bound then the magnetisation is transferred onto the ligand and signals from the hydrogens of the related ligand are detected.[293,309]

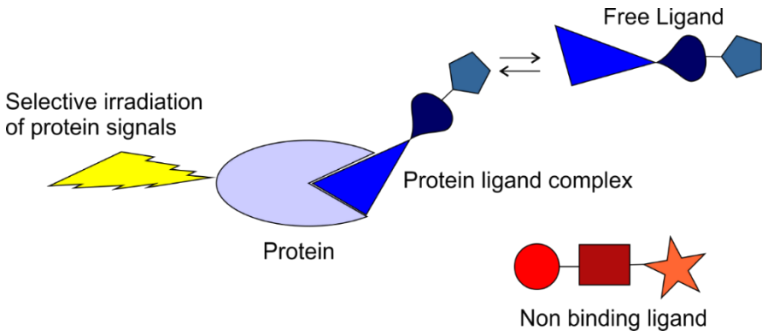


Figure 4.2- A representation of the STD experiment (adapted from [291]). A protein is irradiated at a selective position where no ligand signal exists, the protein becomes saturated and the bulk magnetisation is transferred to the bound ligand (protein ligand complex). The ligand is in exchange between the bound and free form, the saturation is then transferred to the ligand. If there is binding he resultant NMR spectrum contains the signals from protons in the ligand that are in close proximity to the protein. This method can therefore be used to indicate which part of a small molecule forms the binding interface with the protein.

4.1.2 Examples of Fragment-based drug development

A recent review by Turnbull *et al.* details some of the success stories and the techniques involved in fragment-based drug development.[310]

One example is the bromo domain BRD4. This protein has been implicated as a possible anti-cancer target and Zhao *et al.* describe the development of a fragment-based library which was used to screen against this molecular target.[311] The library was composed of fragments from the online database ZINCPharmer; which were selected through the application of a number of filters including “a molecular weight ≤ 250 Da, number of rotatable bonds ≤ 5, logP ≤3.5 and 1≤ smallest set of smallest ring ≤4”.[92,311]

The computational filters were applied using the programme Pipeline Pilot, the fragments were then clustered according to their molecular structure and the central compound of each cluster was selected.[125,311] From this computational protocol 487 compounds were selected and their binding interactions with the BRD4 domain predicted using the molecular docking system Glide.[311,312] The molecular docking elucidated 41 hit fragments which were subsequently crystallised with BRD4 domain.[311] The results of the crystal trials (Figure 4.3) resulted in 9 fragment leads including **4.1 - 4.4,** that were taken forward through further optimisation processes.[311] This research is an excellent example of the use of computational methods to develop compound libraries for screening purposes, it was published after the design and implementation of the approach used in this chapter which is very similar in its nature.

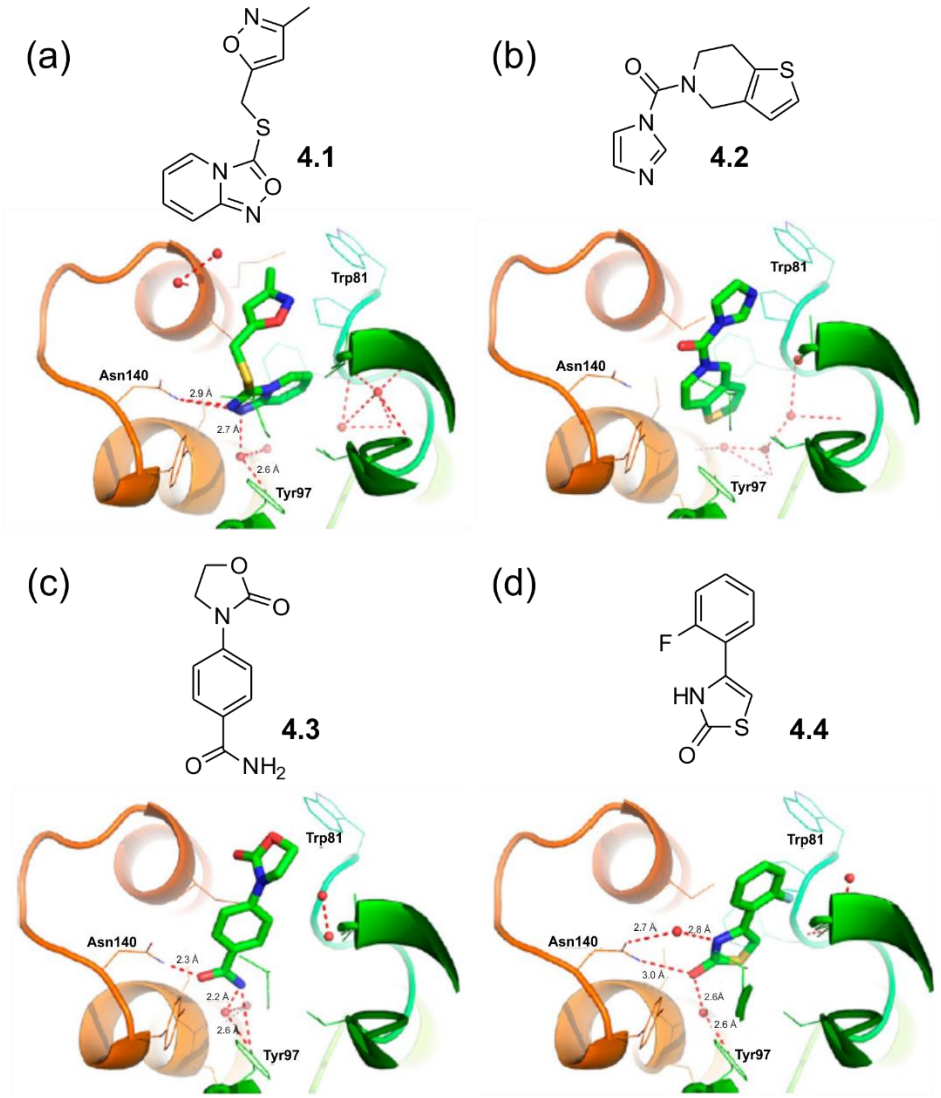


Figure 4.3- Co-crystal structures of the BRD4 domain with four fragments **(4.1 - 4.4)** developed by Zhao *et al.* **(a)** Fragment **4.1** PDB code 4HXO. **(b)** Fragment **4.2** PDB code 4HXK. **(c)** Fragment **4.3** PDB code 4HXP. **(d)** Fragment **4.4** PDB code 4HXN (adapted from [311]).

Other examples of protein-protein interactions that have been successfully targeted using a fragment-based screening approach include the BRCA2 - RAD51 interaction and Prostaglandin D synthase.

The BRCA2 - RAD51 interaction occurs through eight repeating units of the BRC protein, the interaction is essential for DNA repair and its disruption could be employed to sensitise cancerous cells to DNA damaging agents.[313] The identification of the crystal structure of the BRC unit BRC4 in complex with RAD51, elucidated a hotspot motif F-X-X-A (Figure 4.4).[314] The phenylalanine (F) separated from the alanine (A) by two non-important amino acid residues, can be found in small well defined pockets that are able to encompass the side chains of the phenylalanine and alanine residues.[314]

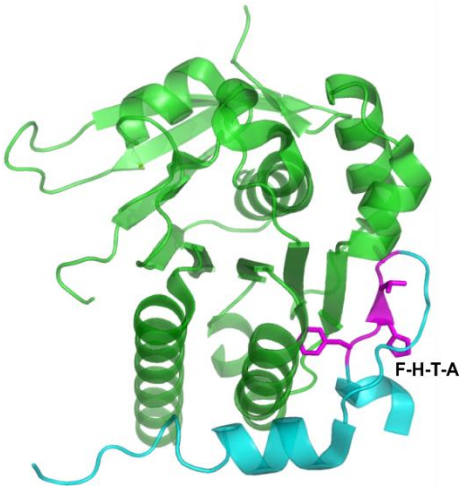


Figure 4.4– A cartoon representation of the crystal structure of the BRCA2 - RAD51 interaction with the F-X-X-A motif highlighted in magenta with the side chains represented.(PDB 1N0W) [314]

The fragment-based approach used to target this interaction employed an initial thermal shift assay screen of 1249 fragments, from which two hits, both with a central indole scaffold, were identified (**4.5** and **4.6**).[314] Binding affinity studies using ITC determined both compounds as having a K_D of 2.1 mM and crystal trials elucidated that they both bound within the pocket of the F-X-X-A motif. Using the central indole scaffold, 42 compounds were designed to investigate the SAR further.[314] To do this Scott *et al.* utilised competitive 1-D ^{1}H STD experiments of the 42 ligands, where a compound known to undergo a binding interaction with the protein, 5-hydroxyindole was added to each NMR sample of protein and ligand. The STD signals of 5-hydroxyindole were monitored to observe any changes in intensity. This was followed by the screening of another 120 fragments that were selected through *in silico* screening and evaluation of the commercial availability.[314] These investigations resulted in six fragment hit compounds for further development (including **4.7, 4.8** and **4.9**), which bound to the target region with an improved affinity in the region of 430, 460 and 570 μM respectively.[314]

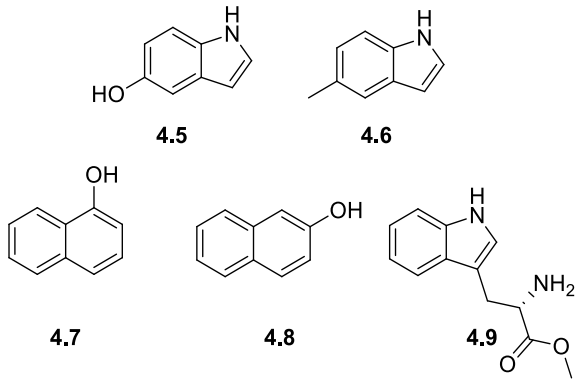


Figure 4.5– A selection of the hit fragments developed by Zhao et al. to target the BRCA2 - RAD51 interaction. The two initial indole scaffolds (**4.5** and **4.6**) from which the fragments (**4.7 - 4.9**) were developed (adapted from [310]).

Figure 4.6- A cartoon representation of the human (H) PGDS dimer (PDB 2VCQ). [315]

Prostaglandin D synthase (PGDS) has been implicated in numerous physiological disorders such as asthma and as such is of interest for therapeutic development.[315] Hohwy *et al.* at AstraZeneca took two fragment libraries of 2450 compounds; the first was composed of 2000 non-specific fragments and the second contained 450 compounds designed to be specific to the PGDS binding pocket.[315]

The second PGDS designed library, was obtained from 20,670 "in-house" compounds, which underwent a series of computational processes.[315] The first involved filtering to select those compounds with a molecular weight < 200 and c log P ≤ 2.[315] The hit compounds were then docked into PGDS using the molecular docking programme GOLD and those compounds that were not found to undergo a hydrogen bonding interaction with the key residues Arg14 or Try152 were removed.[99,315] This process resulted in 450 compounds that were used for biophysical screening.[315] Their primary NMR screen involved testing drug cocktails containing 12 fragments (each at a concentration of 400 μM) against PGDS.[315] The techniques employed included both 1-D waterLOGSY experiments and 2-D HSQC experiments.[315] This NMR based initial screen elucidated 24 possible hit compounds, six of which were found to be from the target specific library of compounds.[315] The fragments were subsequently crystallised with PGDS and molecular docking simulations lead to a further selection of compounds from the "in-house" collection.[315] A second round of NMR screening was then used in conjunction with further *in vitro* assays, to identify the best binders and refine the SAR.[315] Resulting in the identification of four lead chemical series (**4.10 - 4.13**) and an inhibitor (**4.10**) with an IC_{50} =21 nM.[315]

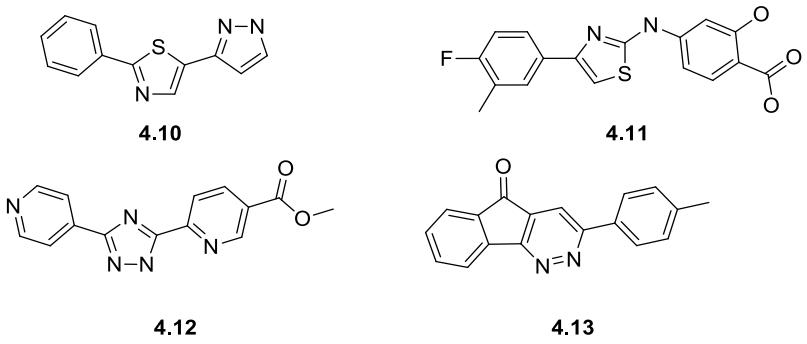


Figure 4.7- Compounds from the four hit series developed by Hohwy *et al*. Developed to target PGDS elucidated through a fragment-based approach, resulting in a lead compound **(4.6)** with an IC_{50} of 21 nM. (adapted from [315])

These examples highlight the combined use of computational design processes and biophysical analysis in the fragment screening of compounds to target different protein-protein interactions.

4.1.2 Aims of Chapter

As previously discussed in Chapter 1, a targeted methodology was developed and used in this thesis to design five small molecules that would perturb the protein-protein interaction between NCS1 and the D2R peptide (Chapter 2 and Chapter 3). The initial screening section of this method involved applying a computational pipeline to a library of general chemical compounds, and lead to the selection of a number of small molecules for synthesis.

This chapter describes the application of the same computational methodology on a library of small molecules that has been designed to be rich in compounds which have been found to perturb a range of different protein-protein interactions. The chemical structures for this library was provided courtesy of Professor Raymond Norton, Monash institute of pharmaceutical sciences, Monash University, Melbourne Australia. The original computational pipeline was adapted to initially focus on the solubility of the small molecules as this is often a limiting factor in early stages of drug design and development. The previous examples of fragment-based studies have tended to involve a large amount of x-ray crystallographic data, however the biophysical screen developed for the purposes of this thesis revolves around a combination of NMR spectroscopic experiments.

The exact computational design process developed will be discussed in further detail (Section 4.2), followed by a discussion on the biophysical analysis screen developed and tested, with examples presented relating to a number of small molecules from within the library (Section 4.3).

4.2. Computational Design

The computational approach to designing small molecules to disrupt protein-protein interactions, is a technique that relies on a source of compounds to work. Previous approaches used within this thesis found within Chapter 2.1, virtually screened an online database of commercially available compounds for ligands which match a specific pharmacophore target. This chapter describes a different but complementary approach. It uses the same computational analysis techniques previously described, for a compound library containing 1137 which have been previously refined as being suitable as starting compounds for development as potential protein-protein interaction inhibitors. Utilising a technique similar to that suggested by Christopher Lepre, [296] we focussed the library on a desirable class of compounds, filtered out those which have undesirable ADMET properties and diversify the remaining selection.

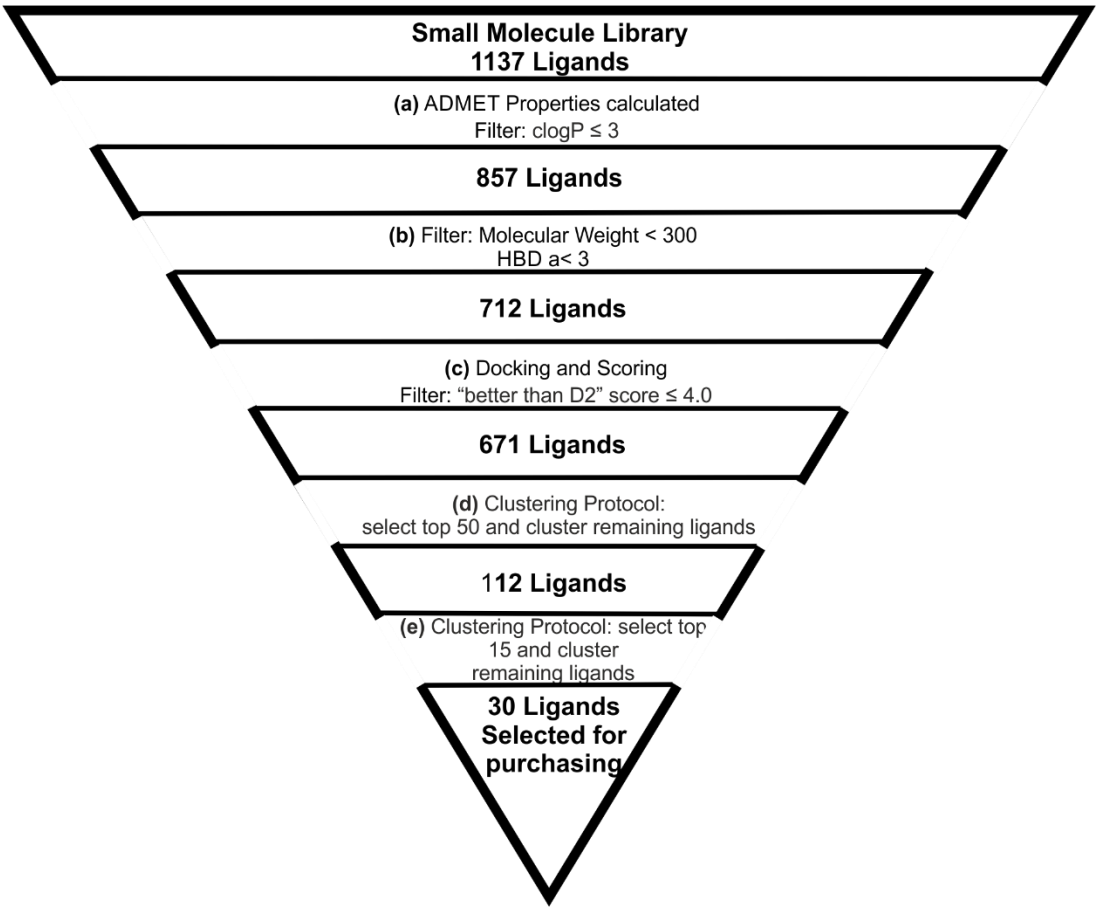


Figure 4.8- The computational workflow applied to the small molecule library of compounds. The initial step in the computational protocol involved filtering those compounds that had a clogP ≤ 3 removing 280 compounds. When a secondary filter is applied to the 857 ligands to remove any that have do not have favourable ADMET parameters such as molecular weight < 300, hydrogen bond donors & hydrogen bond acceptors < 3 removing a further 145 ligands. Molecules were taken forward if they possessed a "better than D2" score ≤ 4.0, this excluded 41 ligands and the final 671 were then ranked in descending order with respects to then "better than D2". The final stages in the pipeline involved the application of a number of clustering procedures, this finally resulted in 30 ligands selected for purchasing.

Due to previous problems encountered with the first generation of inhibitors, it was decided to avoid solubility of the ligand being a major limiting factor in the biophysical screening process, which may result in false positives and synthetic developability issues further down the line. Consequently the first step in the screening protocol involved filtering any compounds that did not have a good solubility profile. To do this, before any further computational screens were applied, the 1137 compounds available in 2-D standard delay format (SDF), were converted to 3-D and energy minimised using modelling programme Spartan '08 V1.2.0 (full details found Chapter 6.1.6.1 Protocol 10).

The second step in the selection process involves the calculation of the ADMET properties (a Figure 4.8 the computational workflow), such as molecular weight and solubility (Figure 4.9 & 6.1.5.3). It is beneficial to calculate these properties for the developability of a fragment library before selecting a candidate compound for testing. Should the ligand be intended for use as an orally available therapeutic agent, it can be used to predict the action of the body on the compound and may help to deselect unsuitable candidates.[113,184] The ease with which a fragment can be optimised is important for the exploration and development of SAR, whilst increasing potency, maintenance of good "drug-like" properties is vital.

With all the chemical data of each compound calculated, a final filter was then applied known as "the rule of three" (b Figure 4.8 the computational workflow).[185] The rule of three is adapted from the original Lipinski's parameters "rule of five" and has been indicated as a useful tool in the development of fragment libraries.[185] As a starting point a work flow was generated as a strategy, with individual nodes used to select those compounds with (a) a molecular weight ≤ 300, (b) clogP ≤ 3 and (c) hydrogen bond donors and hydrogen bond acceptors ≤ 3 respectively being carried forward; those that did not fit the criteria were thus discarded (Chapter 6.1.6.3.2 Protocol 13).[125] Application of this filter removed 425 compounds from the library and resulted in 712 compounds to be taken forward into the next step.

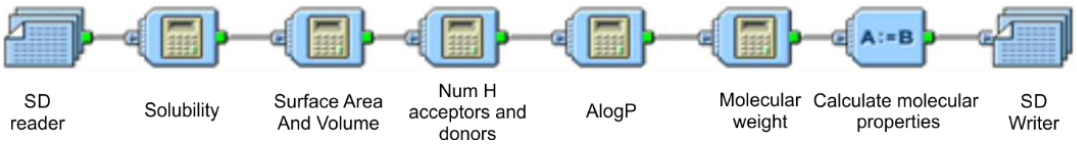


Figure 4.9- Pipeline Pilot workflow used to calculate the ADMET properties of each of the compounds within the compound library, providing a prediction on the action of the body on the compound and may help to deselect unsuitable candidates, should the ligand be intended for use as a therapeutic agent. The files are read into the programme in the SD format, each calculated property (represented by the blue processors) is summarised before the data is written in an output file in the SD format.

With all the ligands in 3-D form the next step involved docking and rescoring them (c, Figure 4.8 the computational workflow). The process of docking involves the prediction of the conformation and orientation of ligand's within the binding site.[238] There are many different docking programmes such as FlexX and Glide; each use different search algorithms and algorithm types. The docking programme used in this thesis was GOLD Suite v5.2, which follows a genetic algorithm that is random or stochastic.[105,238,312] Full details of the processes involved can be found in 6.1.3 molecular docking methods protocols 3 and 4.

Initially the ligands in each file were scored using Goldscore which generated 100 poses in keeping with previous docking protocols employed in the design processes of the first and second generation inhibitors (Chapter 2.1 and 3.1). A fitness value for each pose was generated, based upon the protein-ligand hydrogen bonding energy, protein-ligand Van Der Waals (VDW) energy, the ligand internal VDW energy and the internal torsional strain of the ligand.[99] All files were then rescored in GOLD using three different scoring functions, Astex Statistical Potential (ASP), Chemscore and Chem Piecewise Linear Potential (PLP) (see Chapter 2.1.1 and 2.1.3 for further details of the scoring functions).[99] This combined/ consensus scoring approach was used as each binding score investigates different aspects of the binding process and hence provides a comprehensive account of the predicted ligand binding pose (Figure 4.10). The consensus scoring method is statistically improved, the mean of the samples is closer to a "true" sample than an individual value.[241] The different methods focus on different aspects of ligand binding and so agree on an active compound more than an inactive one as well as providing a greater consistency across receptor systems.[241] The consensus scoring approach has been indicated as leading to improvements in docking procedures, through the improved quality of results with less reliance on the correct selection of a scoring function.[242] This consensus approach to the docking process of the ligands allows for a more informed selection that is not biased, but rather incorporates as many possible aspects of a binding interaction as possible.

To aid the subsequent steps in the computational pipeline the 100 docking poses for each compound were separated. The average Goldscore, ASP, Chemscore and ChemPLP values were calculated from the 100 poses of each individual compound and subsequently merged with the original compound library file (Chapter 6.1.6.3.1 Protocol 12) resulting in a comprehensive file containing all the relevant docking and scoring data.

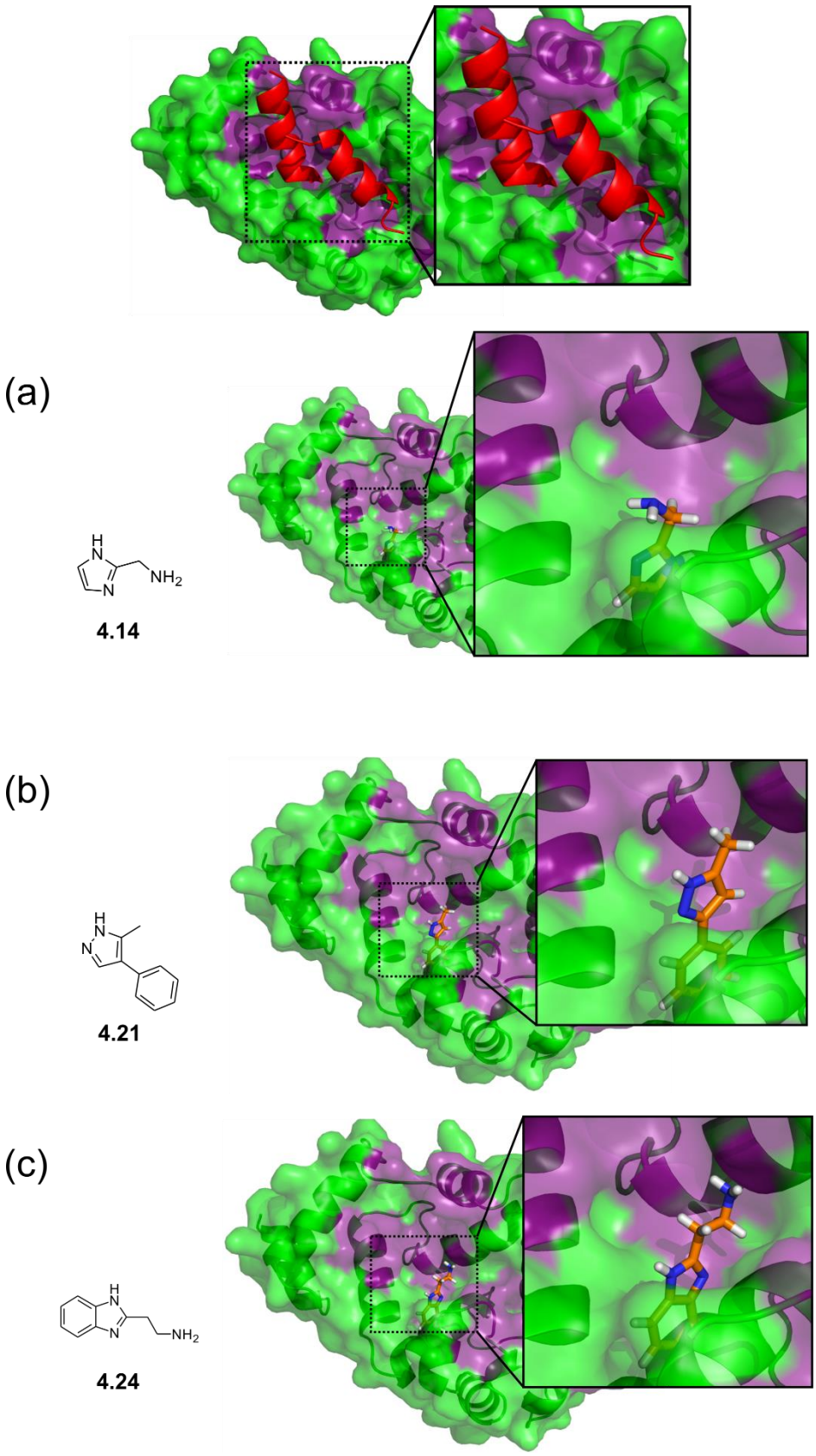


Figure 4.10- The crystal structure of NCS1 (green) and D2R peptide (red), compared to the top poses from the GOLD docking analysis of three ligands from the compound Library with NCS1. **(a) 4.14**, **(b) 4.21** and **(c) 4.24**. NCS1 coloured in green those residues involved with interactions with the D2R peptide are coloured in purple. The top pose for each of the three ligands all appear in a hydrophobic pocket of NCS1 around the EF-2 and EF-3 region.

To process the data, Knime was used (an open-source data integration, processing, analysis, and exploration platform).[124] The four separate scoring files of the 712 compounds (Goldscore, ASP, Chemscore and ChemPLP) were combined and the average ligand efficiencies for each compound was calculated. The ligand efficiencies for each scoring function are calculated based on the average score as a function of the molecular weight (Chapter 6.1.6.3.3 Protocol 14). The calculation for considering whether each docking pose was "Better than D2" was subsequently performed through combining the individual ligand efficiencies for each scoring function of each compound and compared this to that of D2R peptide (equation 4.2). These scores were compared and molecules were taken forward if they possessed a "better than D2" score ≤ 4.0, this excluded 41 ligands and the final 671 were then ranked in descending order with respects to the "better than D2" score.

$$Y = \frac{\varepsilon G}{\varepsilon G1} + \frac{\varepsilon A}{\varepsilon A1} + \frac{\varepsilon C}{\varepsilon C1} + \frac{\varepsilon P}{\varepsilon P1} \qquad 4.2$$

Where Y is the "Better than D2" score and εG, εA, εC, εP represent the four compound "Ligand efficiencies" from "Goldscore", "ASP", "Chemscore" and" ChemPLP" respectively. The values of εG1, εA1, εC1, εP1 are the respective ligand efficiency values for the D2R peptide, calculated from the "Goldscore", "ASP", "Chemscore" and "ChemPLP" results respectively.

With ligands now listed in descending order of the "Better than D2" score, to further the selection process it was deemed necessary to diversify the remaining compounds by clustering them with respects to their "chemical neighbourhood".[316,317] The two successful applications are summarised by (d) and (e) Figure 4.8 the computational workflow, however a more description of the clustering processes shall be discussed in further detail below.

A neighbourhood is defined as the area surrounding the ligand in multidimensional space; this is dictated by the properties of a compound as captured through molecular descriptors such as the molecular fingerprints/ ADMET properties of each ligand.[316,317] In this case the chemical structure of the ligands were used as a "molecular fingerprint".

The distribution of the "Better than D2" score for each ligand was unimodal, with the values ranging from 4.5 - 15.5 and the highest frequency of scores between 8.6 and 9.1 for 14% of the library (Figure 4.11 a). Clustering is based upon the chemical diversity of the compounds; hence in order to cluster molecules the diversity can be specified as a distance between each compound within the clustering protocol

(Chapter 6.1.6.3.4), resulting in 112 separate clusters containing an average number of 6 ligands per cluster. It was initially postulated that for the greatest diversity, the central ligand of each cluster/ neighbourhood should be selected and the distribution of the “Better than D2” values for these compounds investigated, ensuring the chemical space is sampled efficiently and effectively (Figure 4.11 b). However this resulted in a selection of compounds with a lower maximal frequency of the scoring function to between 8.0 & 8.5 for 12.8% of the population. It was decided that this would not be a good method of selection as it reduced the probability of finding a strong binder.

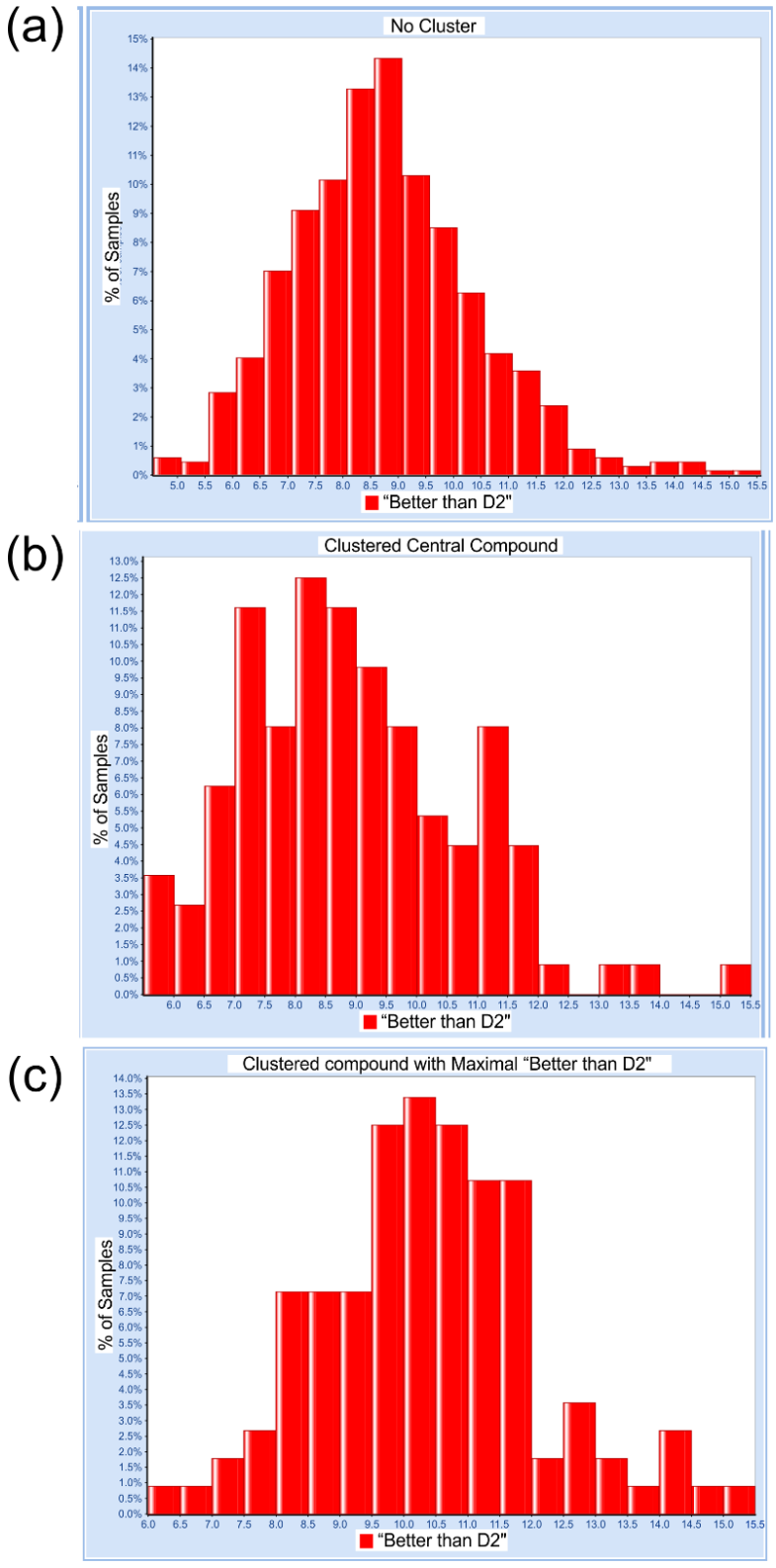


Figure 4.11- Histograms depicting distributions as a result of clustering. **(a)** The distribution of the "Better than D2" score across the library before a cluster is applied, scores range from 4.5 -15.5 with the highest frequency score between 8.6 & 9.1 for around 14.0% of the population. **(b)** The distribution of the "Better than D2" score after application of the cluster and selection of the central molecule within the cluster, the highest frequency score was between 8.0 & 8.5 for around 12.8% of the population. **(c)** The distribution of the "Better than D2" score after application of the cluster and selection of the compound with the top "Better than D2" score, the highest frequency score was between 10 & 10.5 for around 13.0% of the sample.

The next method examined, involved taking the top scoring “Better than D2” ligand for each 112 clusters and assessing the distribution of the “Better than D2” score across the sample. It resulted in uneven distribution of the top “Better than D2” scores with a maximal score of 10 - 10.5 for around 13% of the sample (Figure 4.11 c).

For the most diverse sample selection with the highest probability of a good binding interaction it was decided to select the top 50 compounds with the best “Better than D2” scores and cluster the remaining 620 compounds within the library similar to process employed previously. From each cluster the compound with the best “Better than D2” score was selected, resulting in a sample of 102 ligands from the previous 620.

The distribution of the “Better than D2” scores were analysed and compared to previous results using the same methodology as observed above. It was found that without applying a clustering method to the 102 compounds the most frequent “Better than D2” score was higher than previous. It was observed that 22% of ligands had a score between 10.8 and 11.3 (Figure 4.12 a), compared to the values of the respective sample where only 14% of the population having a “Better than D2” of 8.5 and 9.0 (Figure 4.11 a).

If the 102 compounds were clustered and the central compound of each cluster selected, the distribution of the “Better than D2” score was highest between 11.3 and 11.8 with around 22.5% of the population within this pool (Figure 4.12 b). Compared to the previous corresponding protocol applied to the 620 ligands where the most frequent “Better than D2” was observed between 8.0 and 8.5 for around 12.5% of the population (Figure 4.11 b).

If the clustering was carried out and the compound with the maximal “Better than D2” score selected from each cluster, the distribution of the “Better than D2” score was found to be highest between the values of 11.3 and 11.8 for around 26% of the population (Figure 4.12 c). This appears to be much greater than that of corresponding clustering and selection protocol of the 620 ligands, where we observe the most frequent “Better than D2” to be between 10 and 10.5 for 13.5% of the population (Figure 4.11 c). Also more favourably, the overall distribution of the “Better than D2” scores were much more even and encompassed the higher range of values. Inferring that a better sampling of ligand space had been achieved.

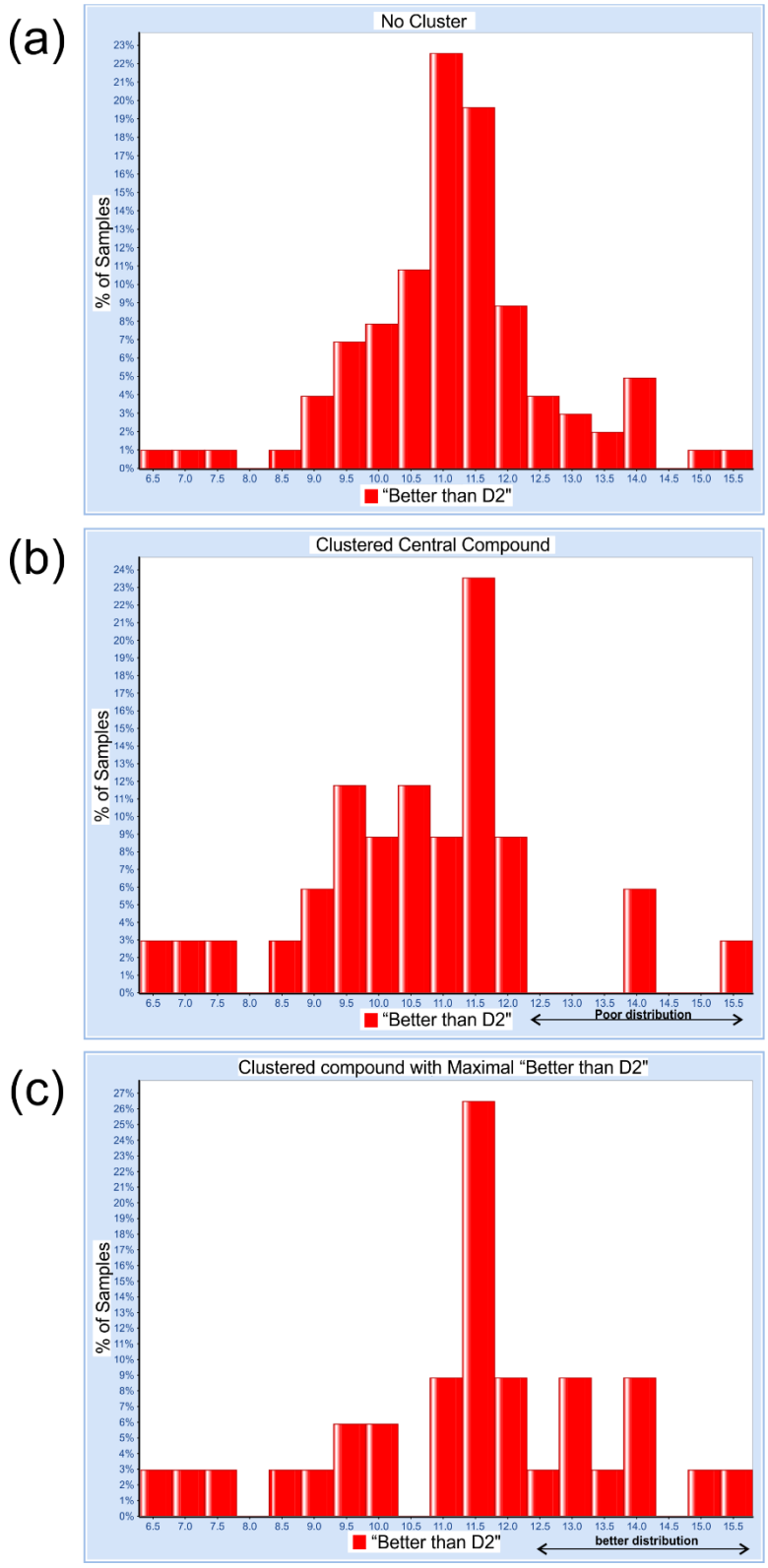
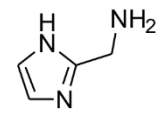
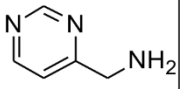
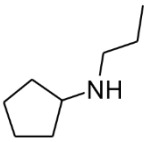
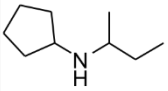
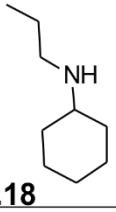
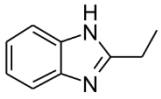
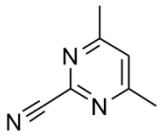
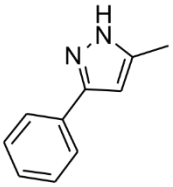
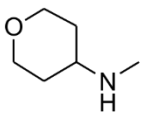
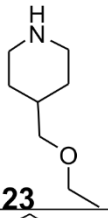
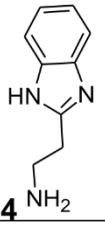
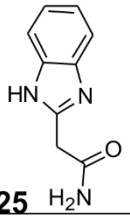
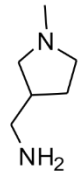
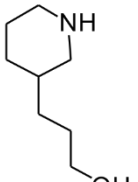
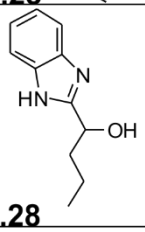
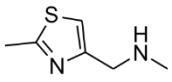
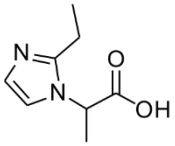
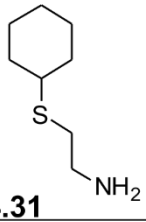
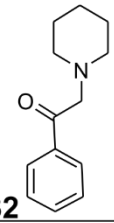
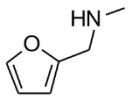
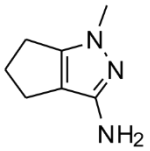
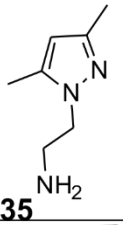
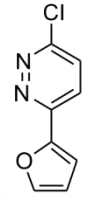
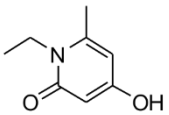
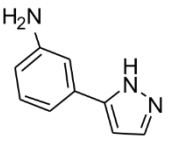
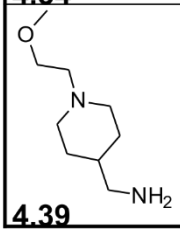
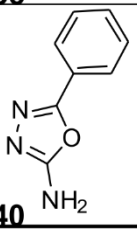
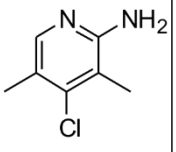
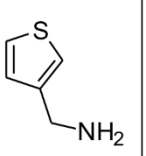
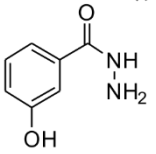


Figure 4.12- Histograms depicting distributions as a result of clustering the 102 compounds. **(a)** The distribution of the "better than D2" score across the library before a cluster is applied, scores range from 6.3 -15.8 with the highest frequency score between 10.8 & 11.3 for around 22.0% of the population. **(b)** The distribution of the "better than D2" score after application of the cluster and selection of the central molecule within the cluster, the highest frequency score was between 11.3 & 11.8 for around 22.5% of the population. **(c)** The distribution of the "better than D2" score after application of the cluster and selection of the compound with the top "Better than D2" score, the highest frequency score was between 11.3 & 11.8 for around 26.0% of the sample.

The distribution of scores in Figure 4.12 c, indicated that this clustering protocol was a good methodology to select the compound sub-library for screening. In order to achieve an economically viable number of compounds (around 30), that our resources would allow, the protocol was repeated. This final selection process took the 102 ligands, the top 15 molecules were selected and the remaining 87 compounds were clustered into 15 clusters. Again the ligand with the best "Better than D2" score from each cluster was selected, affording 30 purchasable compounds for testing.

Table 4.2- Table containing the 30 compounds selected from the screening process of the 1137 within the original compound library. The final two ligands (*****) were unavailable for purchase and so only 28 ligands were tested.

 **4.14**	 **4.15**	 **4.16**	 **4.17**	 **4.18**
 **4.19**	 **4.20**	 **4.21**	 **4.22**	 **4.23**
 **4.24**	 **4.25**	 **4.26**	 **4.27**	 **4.28**
 **4.29**	 **4.30**	 **4.31**	 **4.32**	 **4.33**
 **4.34**	 **4.35**	 **4.36**	 **4.37**	 **4.38**
 **4.39**	 **4.40**	 **4.41**	 * **4.42**	 * **4.43**

4.3. NMR Analysis of Small Molecule Interactions

The computational efforts afforded a targeted compound library, containing 28 small molecules which were subsequently purchased from Life Chemicals-Building blocks© and Chembridge©. The binding interactions of the compound library with NCS1 were investigated using NMR spectroscopy, employing 1-D ^{1}H STD 1-D ^{1}H waterLOGSY and 2-D ^{1}H ^{15}N HSQC experimental techniques; results for each compound within the library can be found in the appendix Section A.5. Each small molecule was dissolved in DMSO achieving a stock concentration of 100 mM (w/v), ^{1}H ^{15}N NCS1 was prepared as described in full in the experimental Section 6.3.2; very briefly the protein was buffer exchanged into the desired NMR buffer and concentrated to achieve a stock concentration of 96.5 µM (m/v). All NMR spectra were collected at 298K unless otherwise stated, on a Bruker AVANCE II+ 600 MHz spectrometer, equipped with 5 mm triple-resonance cryoprobe, the data collected was processed using Topspin 3.1 (Bruker). The NMR samples were prepared to a volume of 550 µL including 10% (v/v) $[^{2}H_2]O$ and the NMR buffer was composed of 50 mM tris-HCl, 50 mM NaCl, 5 mM $CaCl_2$ at pH 6.4.

The experimental screening process followed a two-step protocol;

1) For each compound acquire ^{1}H 1-D, 1-D ^{1}H STD and 1-D ^{1}H waterLOGSY NMR spectra of 1 mM compound in NMR buffer and these will be the reference data. Then acquire the same 1-D data but with a sample of 1 mM ligand and 20 µM ^{1}H ^{15}N NCS1.
2) If the above experiments, termed the primary screen indicated binding between the ligand and protein, then the secondary screen is used: ^{1}H 1-D, 1-D ^{1}H STD, 1-D ^{1}H waterLOGSY and 2-D ^{1}H ^{15}N HSQC experiments. Here the concentration of ^{1}H ^{15}N labelled NCS1 was increased to 50 µM (v/v), making the final Protein: Ligand ratio to be 1: 20.

The 1-D experiments determined which chemical groups of the small molecule were interacting with NCS1. They do not however, provide any information on the protein interaction sites. A 2-D ^{1}H ^{15}N HSQC spectra allow identification of the amino acid residues whose resonances are affected by the presence of the ligand; when mapped onto the structure of the protein, it is often the case that the protein resonances most affected by the presence of the ligand are from residues that reside in and around the binding site and some of these residues will be in direct contacts (hydrogen bond or hydrophobic interactions) with the ligand. All the ^{1}H ^{15}N HSQC were acquired for proteins dissolved in NMR buffer containing 1% DMSO. The “chemical shift

perturbation sum" (CSP) is calculated from the shift differences between the reference spectra of the protein alone (in 1% DMSO) and in the presence of 20 molar equivalent of the ligand (P:L 1:20) and using equation 2.1 from Chapter 2. Where ∆H and ∆N are, respectively, the protein and nitrogen chemical shift changes, and 0.15 is the scaling factor derived from the gyromagnetic ratios of the nitrogen and hydrogen nuclei.[269]

$$\Delta\delta = [\ (\Delta H)^2 + (0.15\ \Delta N)^2\ ]^{1/2} \quad 2.1$$

Histograms of the CSP will contain gaps which can be explained by a lack of assignment for the backbone residue, for example, for residues whose peaks overlap, if the residue has been attenuated or if the peak cannot be determined. Subsequently the sum of total number of residues perturbed was calculated and from this value the top 10% and 20% ranges were determined. The residues whose CSP fell within these two ranges were defined as being within the top 10 and top 20 percentile respectively. This is a direct result of the number amino acids for which assignment has been possible for the individual experiments in question. Therefore making the amino acids of the top 10 and 20 percentile ranges specific to each 2-D ^{1}H ^{15}N HSQC experiment, of each individual compound.

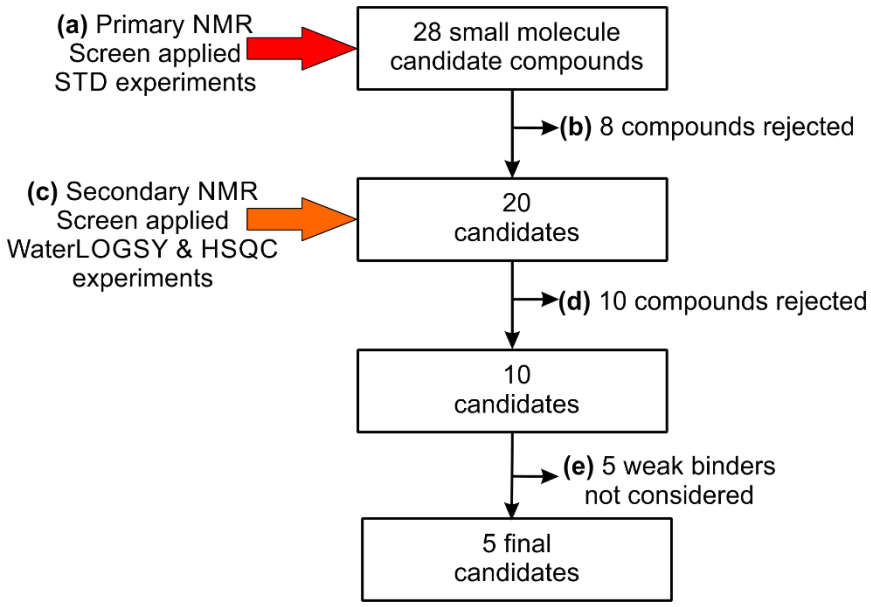
4.3.1 Primary Screen

The results of the entire screening process is summarised in Scheme 4.1; of the 28 compounds the first round of screening experiments which included 1-D ^{1}H, 1-D ^{1}H STD spectra & 1-D ^{1}H waterLOGSY, indicated that eight compounds did not display any form of a possible binding interaction.

The STD data is presented as a difference spectrum; ligand signals that experience a magnetisation transfer from the protein will appear in the opposite phase to that of the irradiated protein resonance. In the case of the 1-D ^{1}H waterLOGSY experiment, binding is indicated by an attenuation of some of the ligand proton resonances.

Figure 4.13, shows the data of two compounds, **4.19** and **4.25**. The waterLOGSY spectrum of the buffer standards are identical to that of the corresponding sample containing 20 µM NCS1; this is the first indication of no interaction between these two compounds and NCS1. The lack of binding is subsequently confirmed with the 1-D ^{1}H STD spectra, where no peaks are likewise observed in the 1-D STD spectrum (Figure 4.13 (black traces)). **4.19** and **4.25**, have a similar chemical structure with a central benzimidazole core (Figure 4.13).

Both the spectra for **4.19** and **4.25** demonstrate that there is no corresponding ligand signals and hence can be concluded that neither ligands are undergoing a binding interaction with NCS1.



Scheme 4.1- The NMR screen process applied to the compound library. The first round of testing rejected 8 compounds of the original 28 as they did not display any precedence for binding to NCS1 in the 1-D ^{1}H STD experiments. 20 candidates of the compound library showed evidence of binding (strong evolution of ligand signals) or a weaker binding interaction (weaker evolution of some but not necessarily all ligand signals). Subsequently the concentration of NCS1 in each of the 20 candidates NMR samples were increased to 50 μM and the previous 1-D ^{1}H, 1-D ^{1}H STD and 1-D ^{1}H waterLOGSY experiments were collected for a second time along with an additional 2-D ^{1}H ^{15}N HSQC spectrum. Of the 20 compounds 10 that were thought to display weak binding were rejected as being false positives as there were no observable changes in the 2-D ^{1}H ^{15}N HSQC spectrum. Leaving 10 candidate compounds, 5 of these were still thought to display weak binding interactions as the changes in the 2-D ^{1}H ^{15}N HSQC spectrum were very minor, this resulted in 5 hit compounds.

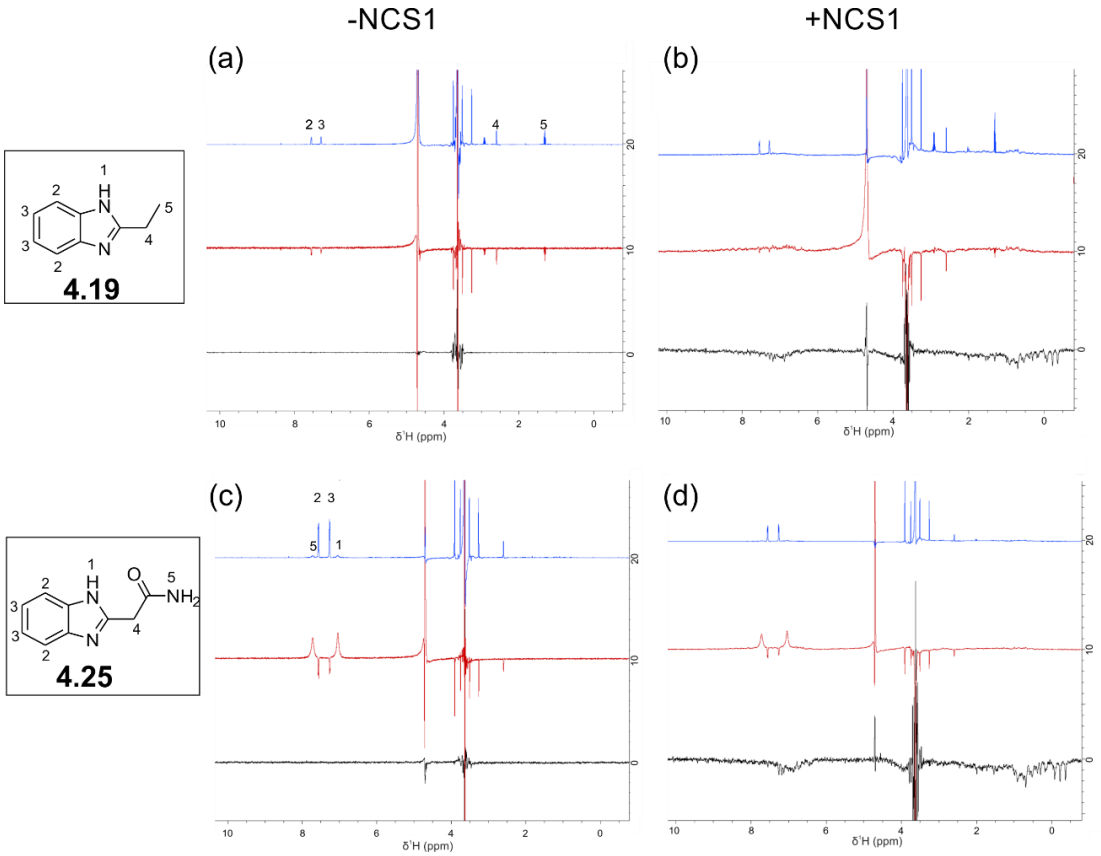


Figure 4.13- Compound library screening data- Two negative compounds from the primary screen applied to the compound library, **(a)** and **(b)** spectral overlays of compound **4.19**, **(c)** and **(d)** spectral overlays of **4.25**. The overlay is composed of the 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black, for both compounds the buffer standard spectral overlay is labelled **(a)** and the protein containing spectra **(b)**. With both ligands there is no apparent difference between the waterLOGSY spectra of the buffer standard and the waterLOGSY spectra of the ligand with the protein, the first indication that there is no interactions occurring between the two. The broad signals in the aromatic region of **4.25** correspond to the rapidly exchanging NH signals from the amide moiety, an artefact of the waterLOGSY experiment. The lack of a binding interaction is then confirmed by the STD spectra, no appearance of ligand signal indicates that there is no transfer of magnetisation between the protein and a bound ligand. (**4.19** ligand signal 1 and **4.25** signal 4 are masked by peaks arising from tris HCl buffer ≈ 3.8 ppm, residual ^{1}H DMSO ≈ 2.9 ppm [318]).

4.3.1. Secondary Screen

The subsequent step in the screening process (Scheme 4.1 c), required the addition of ^{1}H ^{15}N NCS1 to the refined 20 compounds to achieve a final concentration of 50 µM (v/v), and 2-D ^{1}H ^{15}N HSQC spectra were acquired. In addition, the 1-D STD and waterLOGSY experiments are also repeated if a compound from the primary screen was believed to be a weak binder; this step would confirm or reject compounds which were borderline cases in the primary screen. The secondary screen also provided more structural information on the interacting residues of the protein; this along with the data on the interaction with respect to the ligand moieties involved could elucidate a binding pose of the ligand with the protein. Of the 20 compounds that were selected for the secondary screening, 14 compounds displayed a weak binding interaction with NCS1 and 6 compounds showed clearer/ stronger signals in the STD experiments and hence regarded as more promising leads.

An additional compound **4.14** (Figure 4.14) was used as a negative control; it displayed no signs of binding to NCS1 in the primary screening process, a conclusion that was confirmed by the secondary screen. There was no discernible difference between the waterLOGSY spectra of free **4.14** (Figure 4.14 a) and that of the sample containing the protein (Figure 4.14 b). Furthermore, the 2-D ^{1}H ^{15}N HSQC spectra of NCS1 with 1 mM **4.14** show no major changes in the chemical shift protein NH or peak intensities of any of the resonances fond in the spectrum of the control (NCS1 in 1% DMSO). A plot of the combined chemical shift perturbations of the backbone residues of NCS1 with **4.14** (Figure 4.15) indicate that the changes appear small and negligible ranging from 0.00049 - 0.015. Within the overall range of changes in chemical shift, those residues who fall within the top 10 percentile include; Thr17, Gln28, Lys36, Asp37, Asp60, Asn70, Phe72, Phe85, Leu89, Ala104 and Ser184. All but Ser184 are found in the first three EF-hands.

Further analysis of those residues that fall within the 20 percentile include; Val13, Leu16, Thr19, Phe22, Tyr31, Lys32, Gly33, Phe55, Phe56, Asp73, Val180 and Leu189. Again excluding the final two amino acid residues of Val180 and Leu189, the position of the amino acids effected lies within the first two EF-hands of NCS1. As the changes in chemical shift are not of significant ppm value, in comparison for example to that of **4.24** and **4.21**, it can be deduced that **4.14** does not bind to NCS1.

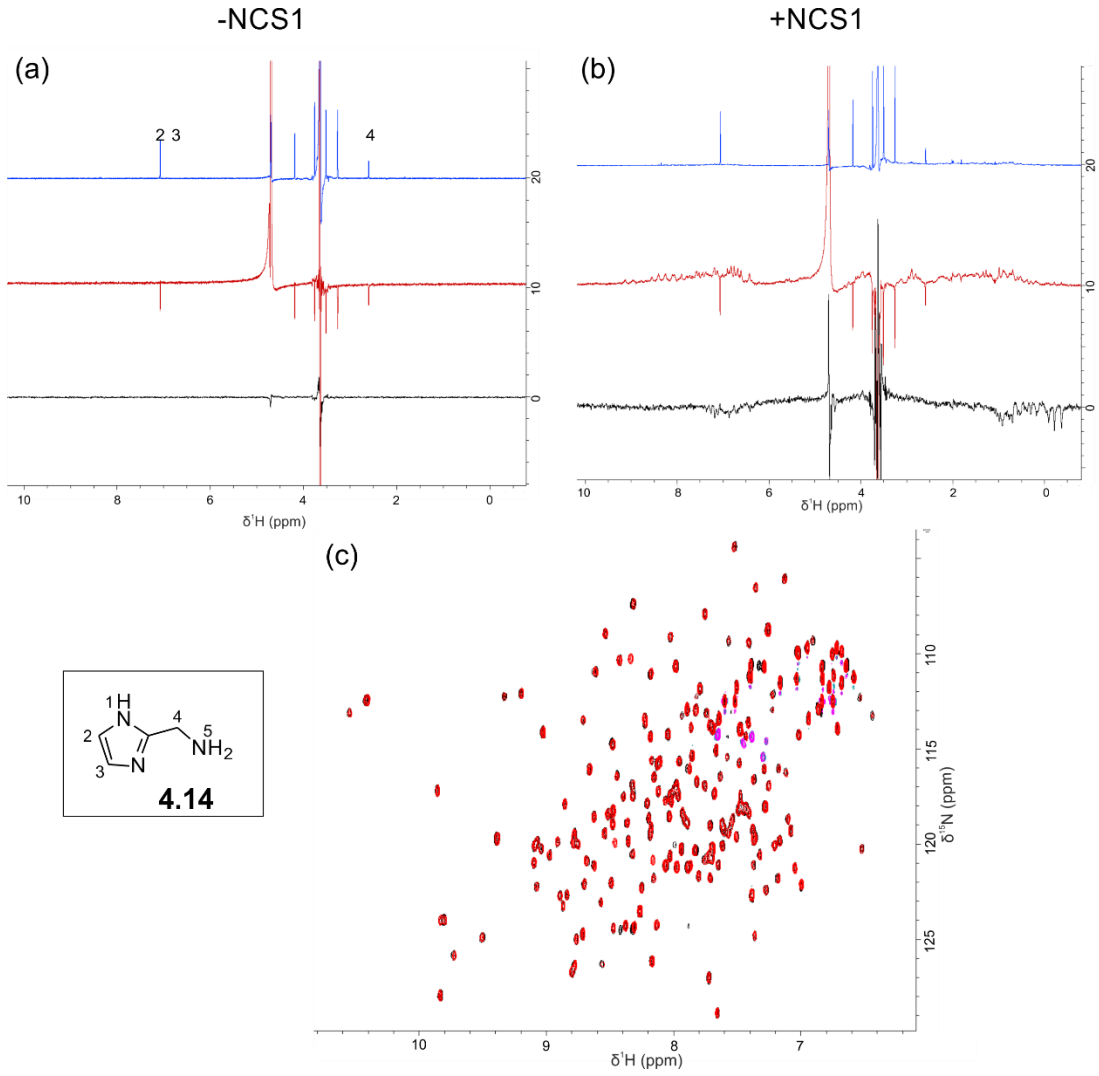


Figure 4.14 -Compound 4.14 NMR screening data. (a) Buffer standard 1 mM **4.14** in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(b)** Binding screen 1 mM **4.14** & NCS1 20 µM in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(c)** 1mM **4.14** & NCS1 20µM in NMR buffer; an overlay of two 2-D ^{1}H ^{15}N HSQC spectra, NCS1 50 µM and NMR buffer + 1% DMSO (black) **4.14** 1 mM in DMSO 1% and NCS1 50 µM (red). Compound **4.14** showed no binding indications in the primary NMR investigations, comparison of the control waterLOGSY spectrum shows no apparent difference, this observation is confirmed in the 1-D ^{1}H STD spectra where the relative ligand signals do not appear, indicating that no transfer of magnetisation is occurring and hence it can be concluded that there is no binding interaction between NCS1 and **4.14**. As a negative control a 2-D ^{1}H ^{15}N HSQC spectrum was acquired an overlay with the NCS1 and DMSO control, no apparent change in the chemical shifts of the amino acid residues of NCS1 further indicate there is no binding interaction.

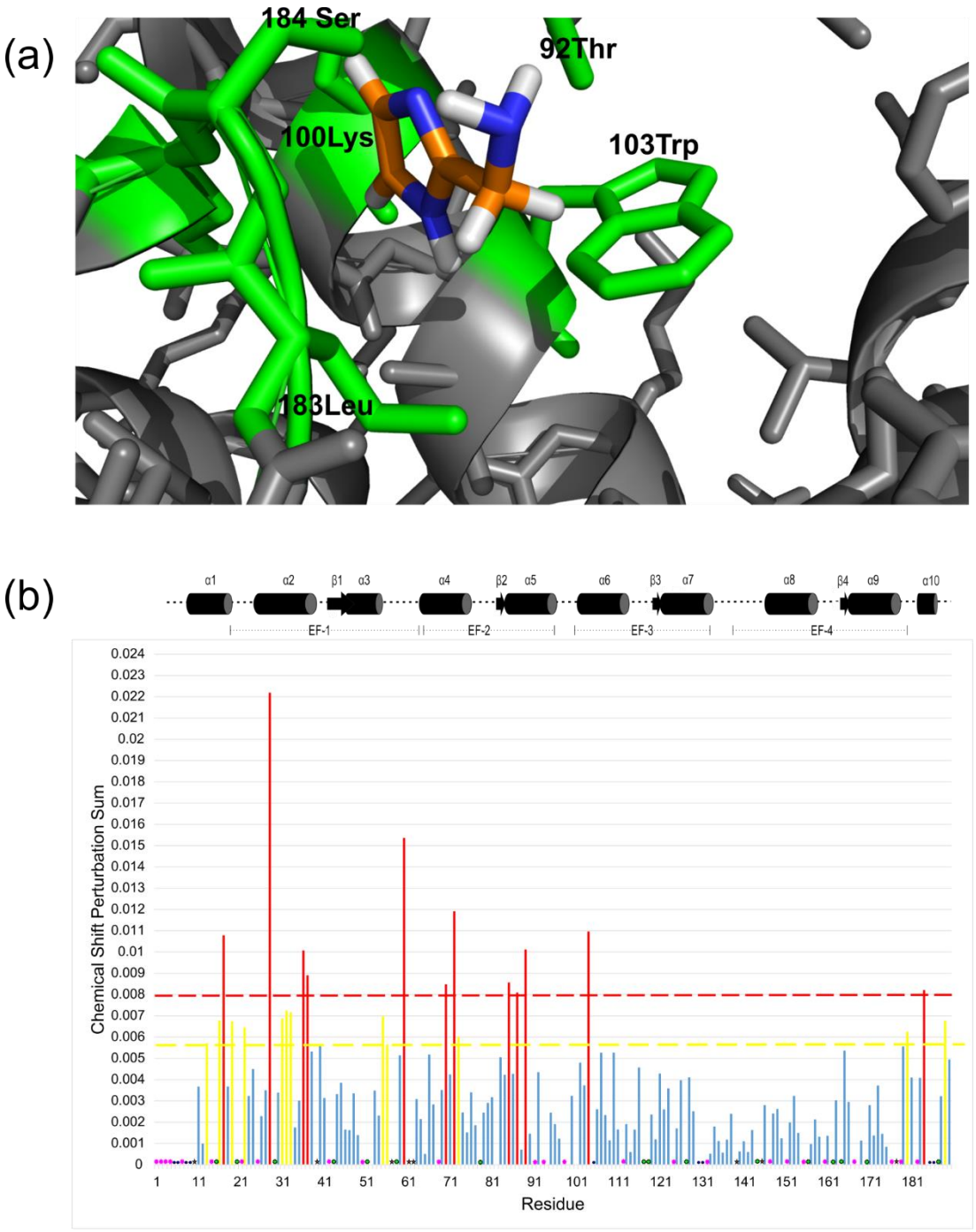


Figure 4.15- (a) Top docking pose of **4.14** in NCS1 (grey) with the side chains of the surrounding amino acids coloured in green and labelled according. **(b)** A plot of the chemical shift pertubation sum differences between free NCS1 with 1% DMSO annd NCS1 with **4.14** and 1% DMSO (Protein:Ligand 1:20). With the protein secondary structure above the histogram. The change in chemical shift of the residues of NCS1 due to the effects of **4.14** appear small and negligible ranging from 0.0005 - 0.02 ppm. Demonstrating that **4.14** does not undergo a binding interaction with NCS1. The red line represents those residues who's change in chemical shift was within the top 10 percentile, the residues that falls within this range are Thr17, Gln28, Lys36, Asp37, Asp60, Asn70, Phe72, Phe85, Leu89, Ala104 and Ser184.. The position of these residues within NCS1 are based within the first two EF-hands α helices 1-5 (α1 - α5). The yellow line line represents those residues that change in chemical shift was within the top 20 percentile, Val13, Leu16, Thr19, Phe22, Tyr31, Lys32, Gly33, Phe55, Phe56, Asp73, Val180 and Leu189. These residues are situated across three of the four EF-hands, EF-1, EF-2 and EF-4, indicating no specificity to one particular location. Gaps within the plots of chemical shift perturbation sum can be explained by: • Unassigned residues. • Single peaks assigned with two residues. ⋆ Proline. • Peaks too become too broad to assign.

The 14 compounds of the original 20 (Scheme 4.1 c), that displayed weaker STD signals indicated a possible binding interaction with NCS1; however further spectral data was required to confirm this. When analysing the STD spectra it is important to remember that the strength of the ligand signal is not directly related to the binding constant.[291] Those compounds with a weaker STD signal may not be binding with a weaker binding constant, they may in fact have a slower off rate (the rate of which the bound ligand exchanges with free ligand in solution). Therefore other spectral data are used along with the STD to deduce the binding mode of the compound.

An example of a compound within the library that was thought to have a weak interaction with NCS1 was **4.24** (Figure 4.16). The STD spectra of the primary screening process contained what appeared to be weak ligand signals relating to the hydrogens of positions 2 and 3 (Figure 4.16 a and b). A comparison between the waterLOGSY experiment of the buffer control and the protein containing sample (Figure 4.16 a and b red spectra) indicate a reduction in the relative signal intensities of the ligand at the corresponding positions 2 and 3 in the aromatic region. A reduction in ligand signal intensity, when compared to the buffer control experiment in which the ligand is having no interactions, indicates that the presence of ligand **4.24** is having an effect on the bulk water of the protein. This is inferred as the experimental design of a waterLOGSY pulse sequence works by saturating the bulk water, it is well documented that water can play a pivotol role in protein-ligand binding interactions.[292][293] In previously documented NMR studies water molecules were found to be present at the binding interface.[306] Some were believed to act like a shell of squeezed water, others involved extensive hydrogen bonded water molecules surrounding the ligand.[306] So by saturating the bulk water with magnetisation if a ligand is undergoing a binding interaction that is mediated by water, then as with the STD experiment the magnetisation will be transferred. However in contrast to the STD experiment this is observed as a reduction in the intensity of the ligand signal.

The use of the STD experiment provides knowledge of which proton environments of the ligand are effected by a possible binding interaction with NCS1, which can be used alonside analysis of a 2-D ^{1}H ^{15}N HSQC spectra and enables the relative amino acids of the protein effected by the ligand to be deduced. To do this the 2-D ^{1}H ^{15}N HSQC spectra of NCS1 with the ligand is overlaid with a 2-D ^{1}H ^{15}N HSQC spectra of NCS1 (50 µM) and with 1% DMSO, any changes in peak intensity or perturbations in chemical shift of a residue indicative of an interaction between the ligand and NCS1.

With **4.24** the chemical shift perturbation sum and changes in peak intensity are again very subtle (Figure 4.17). Changes in chemical shifts were larger than those seen with the effects of **4.14**, with values ranging from 0.001 - 0.035. The residues that undergo changes in chemical shift greater than 0.015 are within the top 10 percentile. They include Gln28, Lys36, Asp37, Cys38, Gln54, Asn70, Glu81, Ala104, Tyr108, Gln181, Leu183 and Gly188. The residues most severely effected appear to be situated in two clusters, those around α2 on EF-1 and α9 EF-4. Furthermore the residues that are found to undergo a change between 0.011 - 0.015 are defined as being within the 20 percentile range. Those residues include; Ile35, Gly41, Asp44, Phe55, Leu110, Asp123, Ile128, Gly133, Ser173, Ile179, Val180, Val190. Found to be situated across all four EF-hands these residues indicate the non specific effect of this ligand within this shift range.

This data further indicates that there is a possible weak binding effect occurring between a small number of residues from NCS1 with the ligand **4.24**, which is concordant with the best predicted binding pose from the GOLD docking (Figure 4.17 a). This is observed in a small cluster of residues around Leu183 highlighted in red in the histogram (Figure 4.17 b), which appear to have slightly higher chemical shift perturbations in comparison to the consensus shift changes of the residues of the whole protein (excluding Asp37).

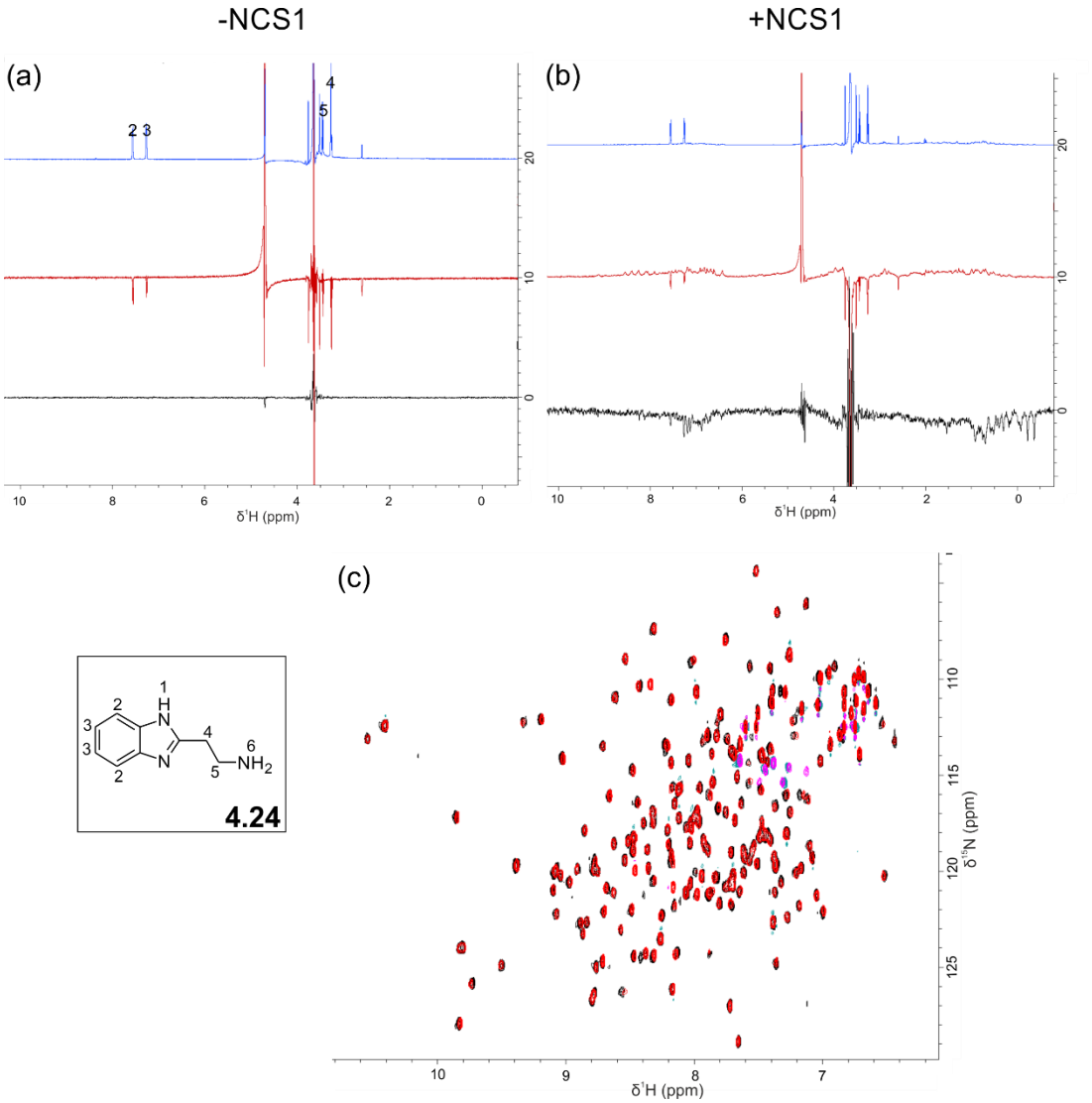


Figure 4.16- Compound 4.24 NMR screening data. (a) Buffer standard 1 mM **4.24** in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(b)** Binding screen 1 mM **4.24** & NCS1 50 µM in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD (black). **(c)** 1 mM **4.24** & NCS1 50 µM in NMR buffer; an overlay of two 2-D ^{1}H ^{15}N HSQC spectra, NCS1 50 µM and DMSO 1% (black) **4.24** 1 mM in DMSO 1% and NCS1 50 µM (red). The results of the primary screen for **4.24** indicated the possibility of a weak binding interaction between the ligand and NCS1. A comparison of the ligand signals 2 and 3 in the waterLOGSY of the negative control (a) with that of the protein containing spectra (b) show a reduction in signal intensity. Secondly, analysis of the STD experiment indicates the weak presence of the ligand signal in the downfield aromatic region. Further investigation into the binding interaction through analysis of the 2-D ^{1}H ^{15}N HSQC spectra (c) is indicative of a weak interaction, with slight perturbations in chemical shift and reductions in the intensities of some signals, for example some changes belong to residues present on the α helix α3 and α5, β-sheet β1 and β2 and in the terminal flexible linker region. This data also indicates that **4.24** does not bind in a specific region.

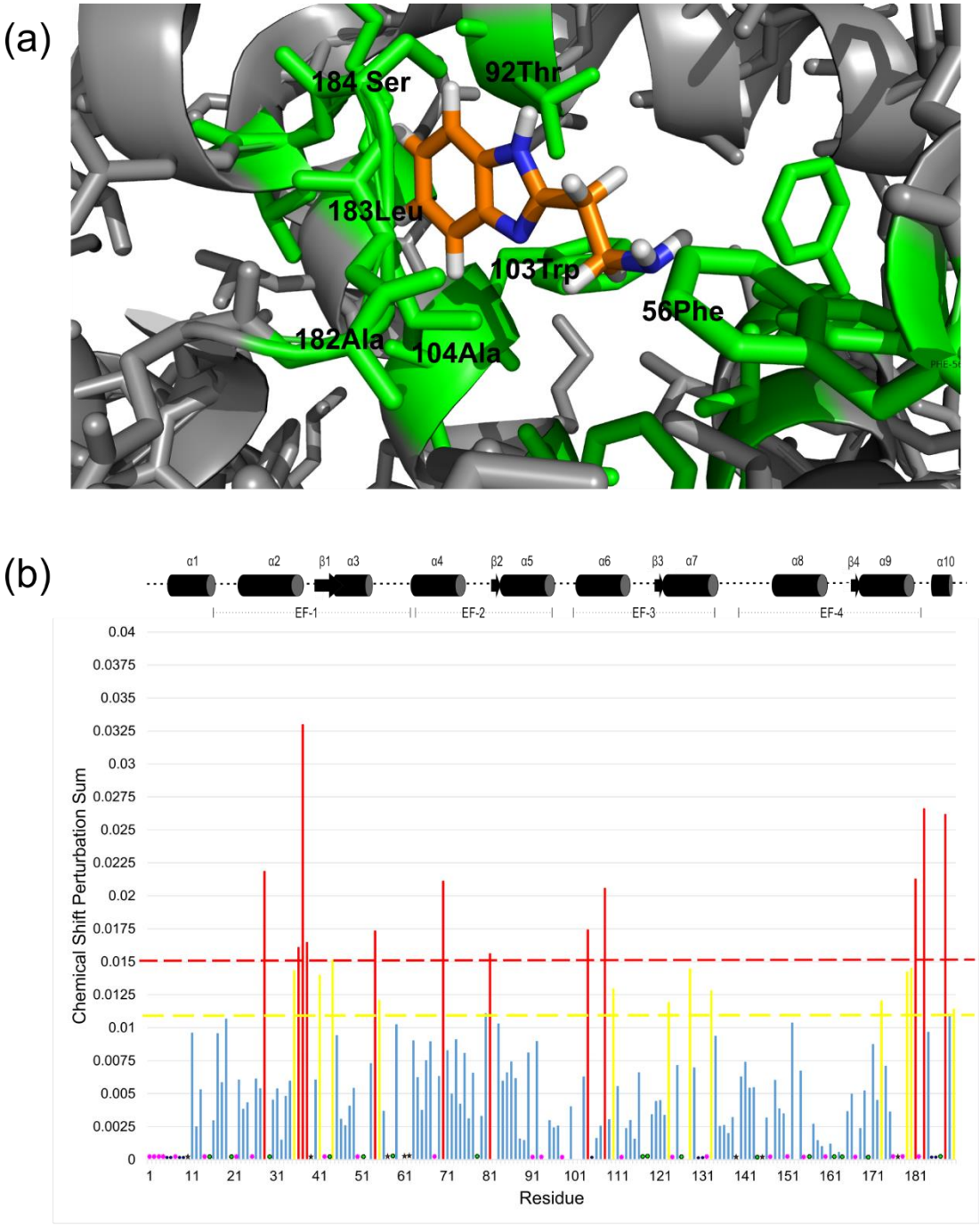


Figure 4.17- (a) Top docking pose of **4.24** in NCS1 (grey) with the side chains of the surrounding amino acids coloured in green and labelled according. **(b)** A full plot of the chemical shift pertubation sum differences between free NCS1 with 1% DMSO annd NCS1 with **4.24** and 1% DMSO (P:L 1:20). The protein secondary structure shown above the histogram. The red line represents those residues who's change in chemical shift was within the top 10 percentile, the residues that fall within this range are include Gln28, Lys36, Asp37, Cys38, Gln54, Asn70, Glu81, Ala104, Tyr108, Gln181, Leu183 and Gly188. The residues most severely effected appear to be situated in two clusters, those around α2 on EF-1 and α9 EF-4. The yellow line line represents those residues who's change in chemical shift was within the top 20 percentile, Ile35, Gly41, Asp44, Phe55, Leu110, Asp123, Ile128, Gly133, Ser173, Ile179, Val180, Val190. Found to be situated across all four EF-hands these residues indicate the non specific effect of this ligand within this shift range. Gaps within the plots of chemical shift perturbation sum can be explained by: • Unassigned residues. • Single peaks assigned with two residues. ⋆ Proline. • Peaks too become too broad to assign.

As described previously, of the 20 compounds that were selected for the secondary screening (scheme 4.1 c), 14 compounds displayed a weak binding interaction with NCS1 and 6 compounds showed clearer/ stronger signals in the STD experiments and hence regarded as more promising leads. Of the 14 possible weak binding ligands only four were shown to have an effect on NCS1 through the secondary screening process therefore, 10 ligands were rejected as false positives (scheme 4.1 d).

The process above had further refined the library, of those 10 remaining compounds it was found that five have only weak effects as monitored by the waterLOGSY and HSQC methods (scheme 4.1 e). This left five possible hit ligands. These five final ligand candidates were first marked as being potential hits in the primary screen, this was subsequently confirmed in the secondary round of testing, the 5 compounds are **4.21**, **4.28**, **4.36**, **4.40** and **4.41** (see appendix A.5 for all experimental results). One of the most promising of the five ligands is **4.21,** a phenyl pyrazole (Figure 4.18); when analysing the data an obvious observation is the complete attenuation of the ligand resonances 1, 2, 4, 5 & 6 in the waterLOGSY spectrum for the protein containing sample (Figure 4.18 b); a second observation from this spectra is the inversion of the methyl ligand signal 3. This is in stark contrast to the signals observed for **4.21** in the buffer standard experiment (Figure 4.18 (a)) where all ligand NMR signals are observable and are of the opposite phase to the protein. When we overlay the waterLOGSY spectra with that of the STD, we notice that the corresponding peaks of the ligand are very much present and that their intensity is very strong. The observable intensity of the ligand signals within the STD spectra is not directly related to the binding constant of the interaction, it is related to the rate at which the ligand is able to disassociate from the protein complex and transfer the bulk saturation to the free ligand in solution. But it can also be inferred that if a ligand has a weaker binding interaction then the probability of it being bound to the protein is less, hence less magnetisation can be transferred resulting in weak signal.[291]

To once again validate the STD and waterLOGSY data of the primary screen, the interaction was further investigated through the use of a 2-D ^{1}H ^{15}N HSQC experiment. When overlaid with a negative control of NCS1 with DMSO, binding interactions were indicated through the change in chemical shifts of a number of amino acid residues, some of which are present on the α helices α2, α3, α6 and α9, β-sheet β1 and β2 and four separate flexible linker regions. A number of the residues affected are located within the hydrophobic binding pocket.

The plot of the chemical shift perturbation sum (Figure 4.19), indicates that the change in chemical shift brought about by the effect of **4.21** on NCS1 occurs over a range of 0.003 - 0.09. This is a substantial increase in the change in chemical shift when compared to the effects caused by both **4.24** and negative control **4.14**. The top 10 percentile are found above 0.065, the residues that fall within this range are Cys38, Asp44, Gln54, Asn70, Thr92, Tyr108, Val125, Ile128, Gly133, Gln181, Leu183 and Gly188. The positions of these residues cannot be located to one single region on NCS1; rather they occur over all four EF hands but do not occur on α1 or α8. In addition to this, the residues whose CSP is greater than 0.048 but less than 0.065 are defined as being within the 20 percentle range. Those residues include Asp37, Gly41, Phe56, Gly59, Phe67, Phe69, Ile86, Gln87, Leu110, Phe169, Gly171 and Ser184, once again they can be found over all four EF hands although not in α1 or α8.

Interestingly four of the residues within the 10 percentile threshold are found in the same region of NCS1 as predicted through the GOLD docking analysis, two of which Thr92 and Leu183 are conserved. It should be noted that the chemical shift perturbation sum is derived from the backbone residues it does not take into account the side chains, it is observable in the 2-D ^{1}H ^{15}N HSQC spectra that there is a major change in the chemical shift or the side chain of Trp103, a residue that was also present in the GOLD docking predicted binding pose. Although the amino acid residues implicated in the binding interaction appear to encompass all four EF hands of NCS1, there is not a systemic change in the chemical shift of all the amino acid residues within the protein. Suggesting that there is a binding interaction occurring between **4.21** and NCS1 however, whether or not this effect is the additive effect of more than one **4.21** ligand would need to be determined. To further investigate this conclusion a STD competition experiment with the native ligand D2R peptide was performed.

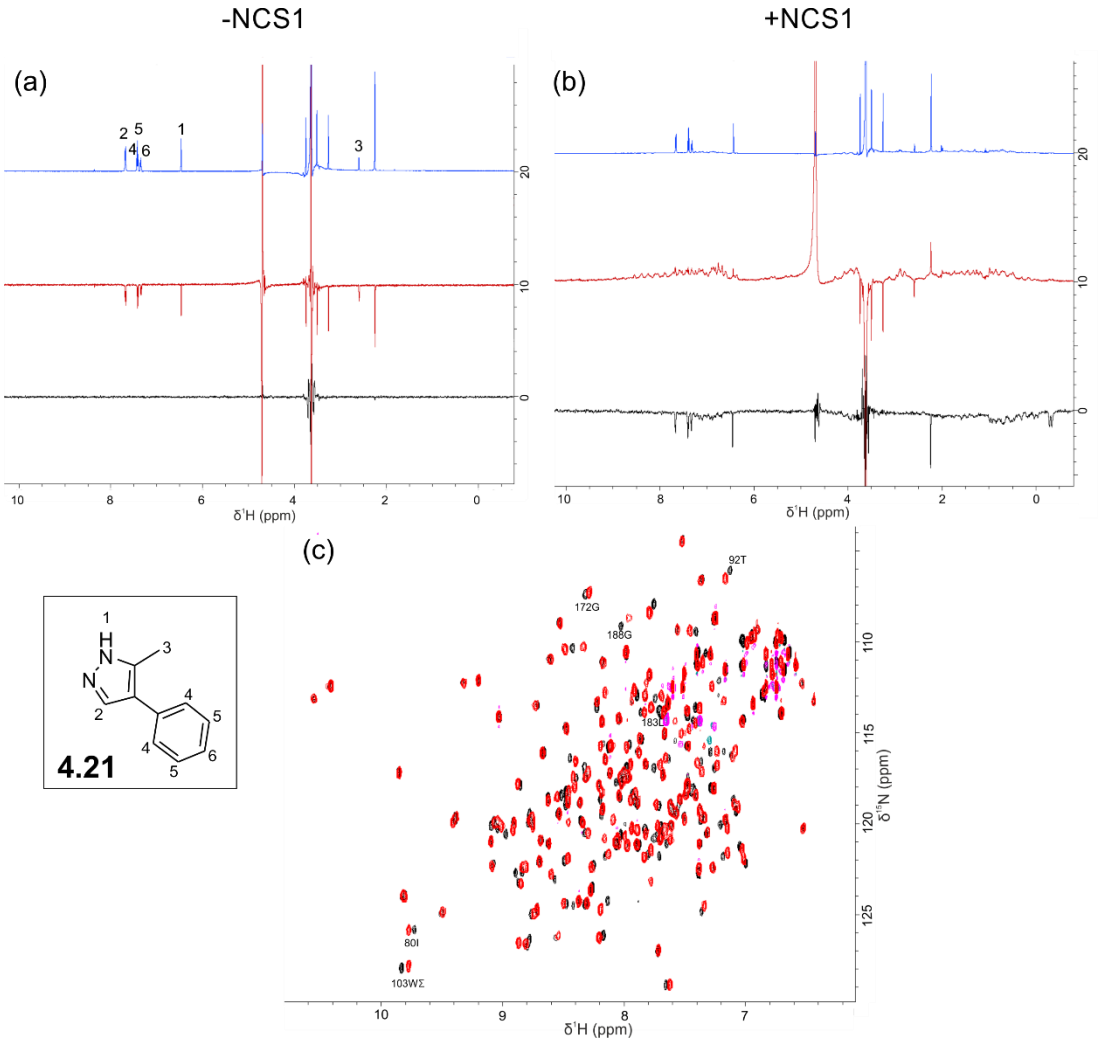


Figure 4.18- Compound 4.21 NMR screening data. **(a)** Buffer standard 1 mM **4.21** in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(b)** Binding screen 1 mM **4.21** & NCS1 50 µM in NMR buffer + 1% DMSO; 1-D ^{1}H spectra (blue), 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD (black). **(c)** 1 mM **4.21** & NCS1 50 µM in NMR buffer + 1% DMSO; an overlay of two 2-D ^{1}H ^{15}N HSQC spectra, NCS1 50 µM and DMSO 1% (black) **4.21** 1 mM in DMSO 1% and NCS1 50 µM (red). The experimental data of **4.21** indicates that the binding interaction between the ligand and NCS1 could be termed as relatively strong. Analysis of the waterLOGSY and STD spectra in the presence of protein (b) are comparable in that the ligand signals 1, 3, 4, 5 & 6 are extremely strong in the STD spectra, suggesting that there is a transfer of magnetisation between the protein and the ligand in this region, also confirmed by the waterLOGSY where the corresponding signals lose all intensity and are completely attenuated. The ligand signal 2 which is also very intense in the STD spectra is then inverted in the corresponding waterLOGSY experiment. This promising interaction was further investigated through the use of a 2-D ^{1}H ^{15}N HSQC experiment which when overlaid with a negative control indicated binding interactions through the perturbations of a number of residues some of which are present a number of which have been labelled, including Ile80, Thr92, TrpΣ103, Gly172, Leu183 & Gly188.

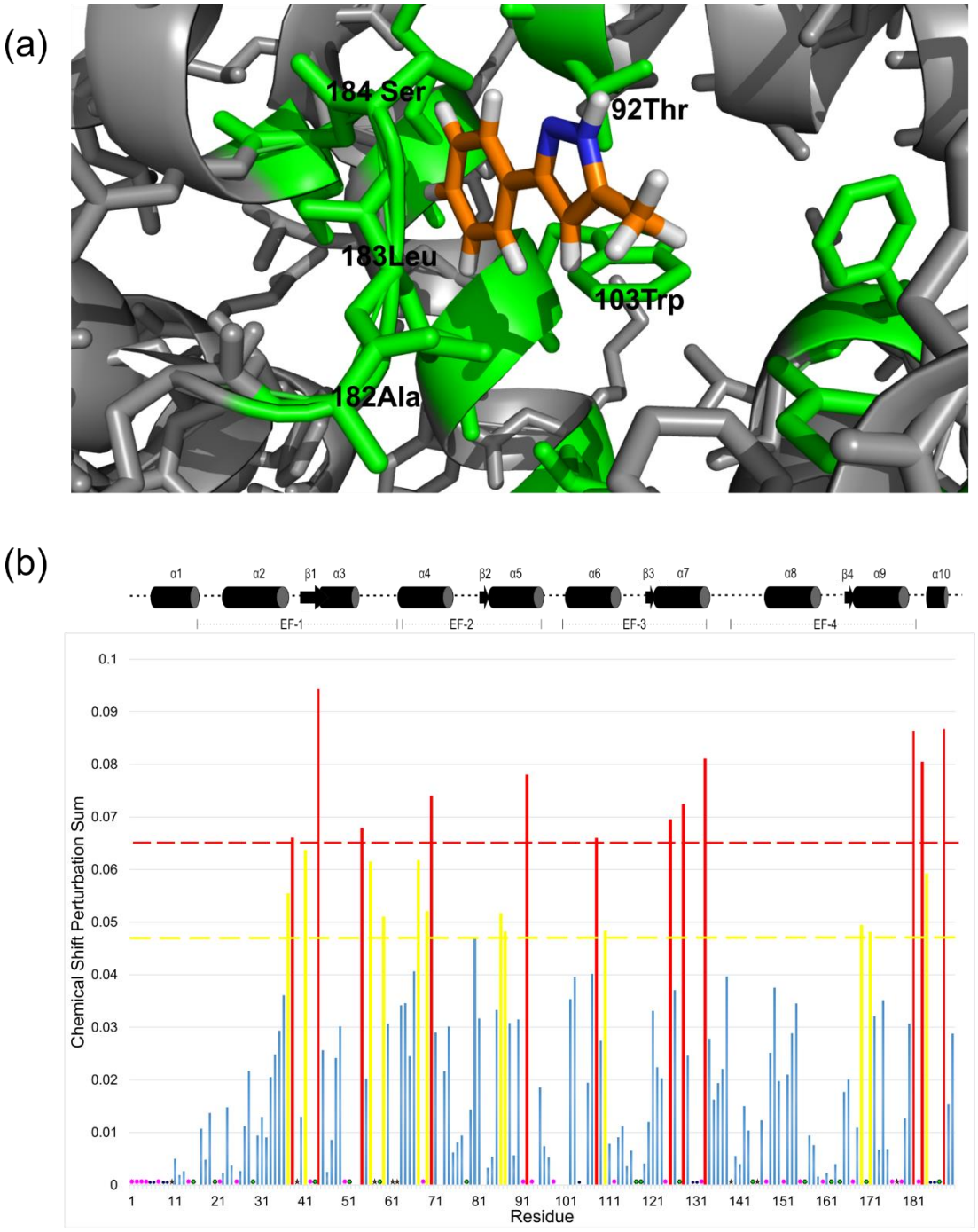


Figure 4.19- (a) Top docking pose of **4.21** in NCS1 (grey) with the side chains of the surrounding amino acids coloured in green and labelled according. **(b)** A plot of the chemical shift pertubation sum differences between free NCS1 with 1% DMSO annd NCS1 with **4.21** and 1% DMSO (P:L 1:20). The protein secondary structure shown above the histogram. The change in chemical shift of the amino acid residues of NCS1 due to the effects of **4.21** occur over a range from 0.0019- 0.12. A large number of residues of NCS1 are affected by **4.21**, the red line represents those residues who's change in chemical shift was within the top 10 percentile. The residues that fall within this range are Cys38, Asp44, Gln54, Asn70, Thr92, Tyr108, Val125, Ile128, Gly133, Gln181, Leu183 and Gly188. The positions of these residues cannot be located to one single regoin on NCS1, they occur over all four EF hands but do not occur on α1 or α8. The yellow line line represents those residues who's change in chemical shift was within the top 20 percentile, and includes Asp37, Gly41, Phe56, Gly59, Phe67, Phe69, Ile86, Gln87, Leu110, Phe169, Gly171 and Ser184. . Gaps within the plots of chemical shift perturbation sum can be explained: • Unassigned residues. ✿ Single peaks assigned with two residues. ★ Proline. ✸ Peaks too become too broad to assign.

To determine if the binding interactions experienced by the 10 ligands including **4.21** were orthosteric or allosteric in nature. To do this the native binding partner D2R peptide, was used in equimolar concentrations to that of NCS1, this is because D2R peptide is known to bind to NCS1 in a 2:1 ligand: peptide ratio with a binding affinity of 30 μM. The results of all 10 compounds used in the competition studies can be found in the appendix A.6 however the results for **4.21** can be seen below in Figure 4.20.

Interestingly upon the addition of D2R peptide there is no apparent change in the ligand signals of the STD spectra, however the intensity of the signals within the waterLOGSY experiment Figure 4.20 (red line), appear increased and much more defined in comparison to that of just the ligand and the protein (Figure 4.18 b). This result could indicate a possible synergistic effect between the D2R peptide and the ligand **4.21**. However problems with the solubility of the D2R peptide in the NMR samples meant that any results from these experiments were inconclusive and cannot be discussed further at this time.

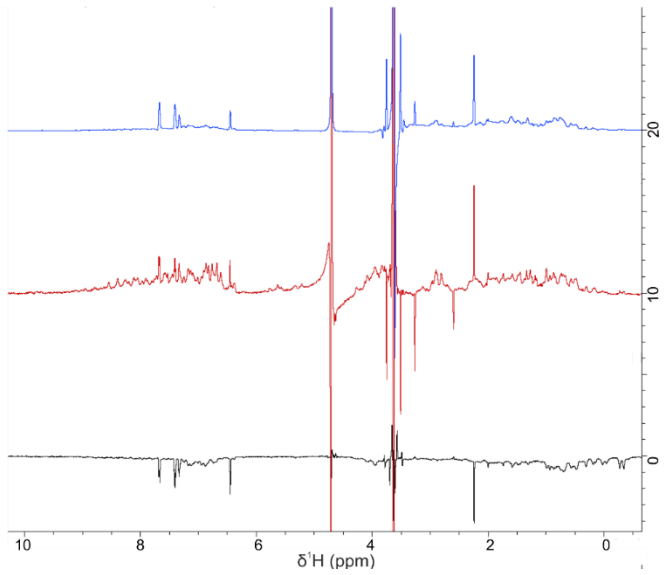


Figure 4.20- Competition experiments of 4.21 with D2R peptide- A 1-D ^{1}H NMR spectra of 50 μM NCS1, 1 mM **4.21** and 50 μM D2R peptide (blue) overlaid with a 1-D ^{1}H waterLOGSY spectra (red) and 1-D ^{1}H STD spectra (black). It is apparent that upon the addition of D2R peptide there is no change in the ligand signals of the STD spectra (black), however the intensity of the signals within the waterLOGSY experiment (red) appear increased and much more defined.

4.3.2. Summary

This chapter has presented an approach to the design of small molecule modulators of protein-protein interactions. It employed a modified version of the computational screening pipeline developed for the purposes of this thesis, adapting the source of the small molecule library. The computational screening was able to focus the library from 1137 ligands to a select number of 28 purchased for screening against NCS1 using biophysical methods.

The biophysical screening process encompassed protein expression, purification and a variety of NMR spectroscopic techniques such as 1-D ^{1}H, 1-D ^{1}H STD and 1-D ^{1}H waterLOGSY & 2-D ^{1}H ^{15}N HSQC. The two step biophysical screening process allowed for the determination of five hit compounds from the original 28, which were found to undergo a binding interaction with NCS1 and five other compounds that had a weaker interaction. Determination of the binding affinities of the five ligands using isothermal titration calorimetry (ITC), and investigations into improving the solubility of the native D2R peptide for competition experiments would be vital for any future work. As well as optimisation studies of the five weaker interacting ligands.

One of the most important aspects of this work, is the proof of principle behind the GOLD docking pose of **4.21** with NCS1. Computational modelling predicted the top binding pose of **4.21** was situated around the residues of Thr92, Trp103, Ala182, Leu183 and Ser184. The experimental results of the chemical shift perturbation sum indicated three of the above residues were conserved between the prediction and the data. Also it is evident through analysis of the HSQC spectra that the chemical shift of the NH side chain belonging to TrpΣ103 is also severely altered. Interestingly **4.21** shares a commonality in chemical structure with that of the second generation hit Inhibitor 5; they both contain a substituted pyrazole moiety, Inhibitor 5 is N-benzylated whereas **4.21** has a phenyl substitution at the 4 position.

Finally this research highlights the beneficial nature of combining two screening techniques, computational modelling and biophysical analysis specifically NMR spectroscopy, in the design of small molecules to target protein-protein interactions. The results of which have led to a promising hit for a small molecule inhibitor of NCS1.

Chapter 5
Conclusions and Future work

5. Conclusions and Future Work

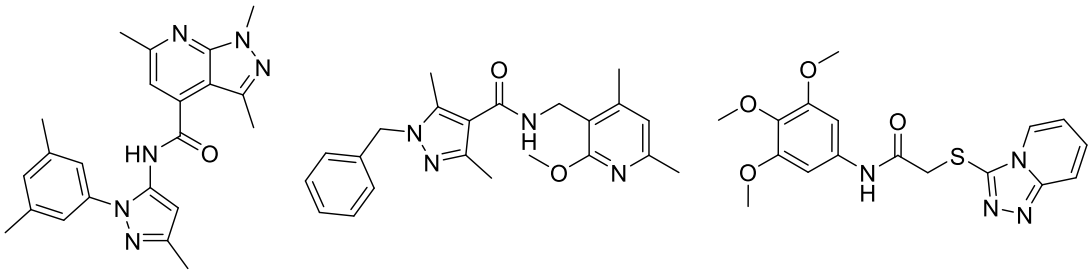
5.1. Targeting NCS1 PPI

Protein-protein interactions (PPIs) are involved in nearly all biological processes and have been implicated in numerous diseases and disorders.[42] Due to their ubiquitous nature and central role, there has been an increasing interest in targeting the interface between two interacting proteins for therapeutic benefit.[42,43] The surfaces involved in PPIs are typically large and flat. As such the molecules which target them tend to have a high molecular weights and are often not traditionally "drug-like" in nature.[43–45,49]

The aim of this research was to discover novel small molecule inhibitors for the interaction between the protein NCS1 and the peptide from the D2 receptor through structure-based drug design. This physiologically relevant interaction has been implicated in such disorders as bi-polar disorder and schizophrenia, areas where novel drug treatments are in high demand (Chapter 1.3.1).[30]

5.1.1 Structure-based drug design

The structure-based drug design process involved the design and implementation of a computational workflow, whereby a pharmacophore-based screening approach of a library of known purchasable compounds was conducted. Optimisation, filtering and validation processes lead to the selection of three of the purchasable compounds which were synthesised (as it was financially unviable to purchase them). The structure-based drug design approach used in this research have never been applied to the NCS1 D2R peptide interaction.



Inhibitor 1 Inhibitor 2 Inhibitor 3

Figure 5.1- The chemical structures of the first generation of inhibitors developed against the NCS1 D2 interaction. Details of the computational workflow, synthesis and biophysical analysis can be found in Chapter 2**.**

The three designed inhibitors of the first generation were, synthesised efficiently in three to four steps, with yields of 28%, 32% and 65% respectively. The interactions of each ligand with NCS1 were investigated using NMR spectroscopy. 2-D ^{1}H ^{15}N

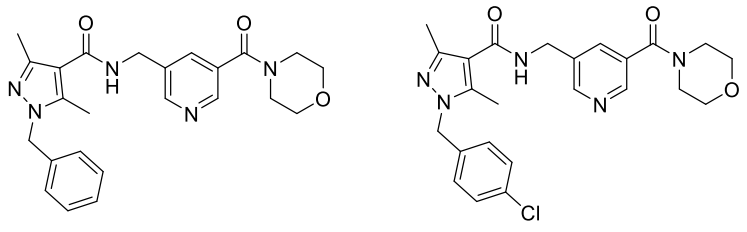
and ^{1}H ^{13}C HSQC experiments showed changes in the amino acids resonances of NCS1, most likely induced through interactions between NCS1 and the molecules. Higher concentrations of Inhibitor 1 were found to cause changes in the spectra indicative of a non-specific interaction (as discussed in Chapter 2.3).

The tryptophan fluorescence results showed a possible change in the folding of the secondary structure of NCS1 in the presence of inhibitors 2 and 3 (due to the observed red and blue shifts in the fluorescence signal), whereas Inhibitor 1 was shown to cause quenching of the fluorescence. To further investigate the binding interactions of the first generation with NCS1, binding affinity was explored using isothermal titration calorimetry (ITC).

Full characterisation of the binding affinity was not possible due to problems associated with the solubility of the inhibitors in aqueous buffer. The solubility issues encountered are a common problem associated with targeting PPIs with small molecules; this is more than likely due to the nature of the interaction sites being hydrophobic.[251,319,320]

The limited solubility of the first generation of inhibitors restricted further biophysical analyses; therefore the logical progression was to develop a second generation of inhibitors with improved solubility to facilitate accurate binding measurements.

In an effort to solubilise the next generation compounds, a variety of solubilising groups were incorporated to the core scaffold of Inhibitor 2 and a number of structural changes were made (Chapter 3.1 Figure 3.2). The computational design efforts and screening pipeline afforded 12 candidate compounds for synthesis. However, challenges encountered during the synthesis (Chapter 3.2) meant that only two of the intended compounds were synthesised; Inhibitors 4 and 5 in yields of 36% and 34% respectively (Figure 5.2).



Inhibitor 4 Inhibitor 5

Figure 5.2- The chemical structures of the second generation of inhibitors developed against the NCS1 D2 interaction. Details of the computational approach, synthesis and biophysical analysis can be found in Chapter 3.

Fluorescence and NMR investigations indicated more specific but weaker binding of the second generation inhibitors compared with first generation. This result was observed in the 2-D ^{1}H ^{15}N and ^{1}H ^{13}C HSQC where Inhibitor 4 appeared to have less of an effect on the amino acid residues of NCS1 than Inhibitor 5. A greater quenching of the intrinsic fluorescence in the presence of Inhibitor 5 than Inhibitor 4 was observed at identical concentrations. Secondly, the concentrations of the second generation required to cause quenching of the tryptophan fluorescence was over 100 times greater than the first generation (as discussed Chapter 2.3.8 Figure 2.48 and Chapter 3.3.3 Figure 3.27).

Despite the improved solubility of the second generation inhibitors, DMSO was still required to dissolve the compounds at the millimolar concentrations required for the ITC experiments. However, because DMSO interacts with NCS1, this method could not be used to obtain more precise K_D values.

The removal of compounds with more unfavourable physicochemical properties, and unfavourable predicted binding poses, meant that the computational protocol was successful in its ability to select compounds with a greater probability of interacting with the target protein. The first and second generation compounds have been shown to undergo an interaction with NCS1, however biophysical analyses were limited due to the solubility challenges the compounds present. These challenges do not discredit the interactions observed between the ligands and the protein in ITC and fluorescence investigations, however it does limit the determination of any quantitative affinity value. To reduce the challenges associated with the use of DMSO and NCS1, a different organic solvent that does not have any effect on the protein should be used to solubilise the compounds for future testing of other ligands.

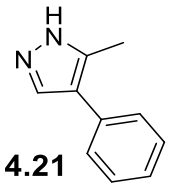
5.1.2 The fragment-based approach

As an alternative to identify potential hit compounds, a fragment-based approach was also carried out. The computational pipeline previously developed, was adapted and applied to a library of small molecule fragments (~1,137), designed to be rich in compounds which have been found to perturb a range of different protein-protein interactions (Professor Raymond Norton, Monash institute of Pharmaceutical Sciences, personal communication).

Due to the solubility restrictions encountered with the structure-based drug design of the first and second generation compounds, the computational pipeline was adapted to circumvent this with the fragment approach. A minimum solubility filter was applied

by application of the “rule of three”, this enabled exclusion of the more insoluble compounds and resulted in a selection of 28 ready to purchase fragments.

A biophysical screening strategy was developed focusing on a pipeline of NMR spectroscopic experiments in order to efficiently screen the 28 compounds and discover those which bind with NCS1. Of the 28 compounds, eight were identified as not interacting with NCS1 through 1-D STD/waterLOGSY NMR experiments (Chapter 4.3.1 Scheme 4.1 a and b). Of the 20 compounds that were identified as interacting with NCS1, 5 underwent a binding interaction with NCS1 and 15 were either false positives or the binding interaction with NCS1 was very weak. Biophysical screening resulted in identification of 5 hit compounds, one of which is an extremely promising lead candidate for the development of inhibitors of the NCS1 D2 interaction due to its greater perturbation of the targeted amino acid residues of NCS1.



4.21

Figure 5.3. The chemical structure of the lead fragment compound **4.21**. Details of the computational workflow applied to the fragment library and biophysical analysis can be found in Chapter 4.

5.1.3 Concluding remarks

We have identified and verified several small molecules that bind NCS1 *via* two differing screening strategies, one which used computational applications to select compounds to synthesise and the other to select fragments to screen. From each approach one molecule has been highlighted as a hit compound for the future development of inhibitors of the NCS1 D2 interaction; Inhibitor 5 and **4.21** (Figure 5.3). Interestingly both compounds share a commonality in their chemical structure; they both contain a substituted pyrazole moiety, Inhibitor 5 is N-benzylated whereas **4.21** has a phenyl substitution at the 4 position. Both approaches to select these compounds involved an initial computational screening protocol, which was key in removing those compounds which were unlikely to bind as wished and those with unfavourable physicochemical properties prior to candidate selection.

The nature of the fragment screening approach made it a much more suitable and efficient method for identifying NCS1 binding compounds and is a method that has been successful in targeting a number of other PPIs.[310] The improved solubility of the smaller fragments in comparison to that of the larger compounds of the first and second generation, improved the handling for biophysical characterisation. From this,

hit fragments then provide a core scaffold which can be further developed into larger compounds able to overcome any non-specific binding interactions that may occur as a result of their smaller size.

This research contributes greatly to the understanding of the challenges associated with targeting the NCS1 D2R peptide interaction with small molecules, specifically highlighting the challenges associated with using organic solvents such DMSO with NCS1. These compounds however, present an exciting initiation point for further investigations into the druggability of this important PPI.

5.2. Future NCS1 research

After reviewing the previous work into the development of inhibitors of PPIs and the research detailed in this thesis, the future research into the development of inhibitors of the NCS1 D2 interaction could include:

Further investigation of the binding interactions of Inhibitor 5 and **4.21** through systematic mutagenesis of the eight key residues of the NCS1 binding groove. This will enable the determination of which portion of the binding site is interacting with the compounds.

Using the structural information collected for all the compounds screened against NCS1 in this thesis, a new chemical scaffold could be designed. This would include an improved solubility profile to allow the investigations of the binding affinity in aqueous buffer without the influences of DMSO. The efficiency and ease of the synthesis should be improved in comparison to that of the second generation; with the aim of generating a higher number of compounds to test. Use of solid matrix lead binding tests such as surface plasmon resonance (SPR) [299] could be used to circumvent any solubility issues that may arise. Finally all structural investigations should be verified through the use of cellular based assays such as ELISA, which have been used previously in investigations of the distribution of NCS1 in the rat brain.[321]

Chapter 6
Experimental

6 Experimental

6.1 Computational Protocols Ligand-based Virtual Screening Methods

6.1.1 Pharmacophore Selection Protocol 1

Selection of the initial pharmacophore of the frequenin (frq1) and phosphatidylinositol 4-kinase (pik1) interaction was carried out informed by previous structural work previously conducted by Ames *et al.* and using PyMOL version 1.3.[34][224] The NMR derived structure of the frq1 pik1 complex was downloaded from the RSCB protein data bank (PDB) (PDB code 2JU0) and loaded into PyMOL.[34] The key hydrophobic residues of pik1 Val-156, Ala-157, Ala-159, Leu160, Val-161, and Met-165 indicated by Ames *et al.* were selected from the PDB file in PyMOL and the subset of the complex saved for pharmacophore screening.[224]

6.1.2 Online Screening Database Protocol 2

Virtual screening of ligand databases for matches to the selected pharmacophore was conducted using the online pharmacophore search software database ZINCPharmer.[92] The pik1 file prepared in PyMOL (6.1.1 Protocol 1) was selected, excluding the hydrogen donors, acceptors and those hydrophobic residues not indicated by Ames *et al.* as being involved in the key interactions. Hit reduction and hit screening filters were then applied to each individual search. For each of the six different pharmacophore's (defined Chapter 2.1) the search filters applied were as follows; molecular weight (Da) ≤ 600, ≤ 500, ≤ 300 and ≤ 150; root mean squared deviation (RMSD) ≤ 2, ≤ 1 and ≤ 0.5 and rotatable bonds ≤ 9 and 9-1, the results for these screens can be seen in Chapter 2.1.2 Table 2.2 – Table 2.7. For each different search the ligand files were downloaded and saved to a directory locally for further analysis.

6.1.3 Molecular Docking Methods

Initially the ligands downloaded for each pharmacophore were docked and scored using the software GOLD Suite v5.1 with the function "Goldscore" (Protocol 3), they were then rescored using Astex statistical potential "ASP", "Chem Score" and Chem Piecewise Linear Potential "ChemPLP" respectively (Protocol 4).[99] Secondary docking studies were conducted upon the ligands interactions with NCS1 (pdb code 5AER) following the same protocols 3 and 4 with amendments to the protein pdb

code. With the NCS1 docking the binding site was determined as 6 Å of “L_1, C:\backup\Helix 1 dopamine.pdb” (one dopamine c terminal helix).[99]

6.1.3.1 Protocol 3

- Load appropriate file of frq1 active site (pdb format).
- Setup and perform docking calculations using “Wizard”.
- “Add all hydrogens” to the protein (1391 in total).
- Define the binding site as 6 Å around the reference position of the pik1 fragment (pik1 residues Ala157-Met167) “L_1Scitegic02071211143D”.
- Select “Detect Cavity”.
- Select “Solvent Accessible.”
- Select ligands saved from protocol 5.1.2.
- Within “Fitness and Search Options” select “Docking” and “Goldscore” as default, select “Allow early termination” and use “Internal ligand energy offset”.
- “Output Options” select “Structure Data (SD) file format”, define output directory and deselect “Save Lone pairs”.
- Submit calculation by selecting “Run Gold” and review results saved automatically in output directory.

6.1.3.2 Protocol 4

- Load appropriate file of frq1 active site (pdb format).
- Select “Gold” tab, “Set up and run a docking”, “Load existing”.
- Select the corresponding previously run Gold docking “conf” file to load the ligands results from the previous docking.
- Keep all settings the same as the default except within “Fitness and Search Options” where “Docking” should be deselected.
- Select “Rescore” and set “Rescoring function” to either “ASP”, “Chem Score” or “ChemPLP”.
- Within “Output Options” define a new output directory for the new file type.
- Submit calculation by selecting “Run Gold” and review results saved automatically in output directory.

6.1.4 Ligand-based Screening and Selection

6.1.4.1 Physiochemical Properties Protocol 5.[125]

Physiochemical properties of the ligands docked in 6.1.3.1 Protocol 3 and 6.1.3.2 Protocol 4 were calculated in Pipeline Pilot Student Edition version 6.1 using the

following components in a sequential protocol (this process was repeated for the NCS1 docking);

- "Chemistry Readers"; "SD Reader". Imports a file in the structural data file format.
- "Calculators"; including "Solubility", "Surface Area and Volume", "Number of Hydrogen Acceptors and Donors", "ALogP, "Molecular Weight" and "Molecular Properties" specifically selected to calculate the number of rotatable bonds. Used to calculate the ADMET properties using a number of components.
- "Chemistry Writers"; "SD Writer". The data files can be extracted and saved in the SD format using for use in the balanced selection process 6.1.4.2 Protocol 6.

6.1.4.2 Balanced Selection Protocol 6.[124]

The balanced selection protocol was conducted in Knime 2.5.1 and the workflow protocol can be seen below.

- SDF Reader: Read in each SDF file for each of the four scoring functions.
- Row Filter: Remove any negative fitness scores.
- Math Formula: Calculate Ligand Efficiency according to the formula:

  Ligand efficiency = Goldscore ÷ molecular weight
- Column Filter: Remove any unwanted or unnecessary information leaving only the following as columns of data- "Molecule Name", "Fitness Score Solubility", "Surface Area and Volume", "Number of Hydrogen Acceptors and Donors", "A LogP", "Molecular Weight", "number of rotatable bonds" and "Ligand efficiency".
- Sorter: Sort the ligands into descending order with respects to the "Ligand efficiency".
- Math Formula: Calculate the average score for the scoring function (ASP, Chemscore, ChemPLP), hence rank each ligand and append as new column.
- Sorter: Re-sort the ligands into ascending order with respects to the row key.
- Joiner: Join two of the ranking files.
- Column Filter: Exclude any duplicate information.
- Joiner: Combine the results from the four different ranking files.
- Column Filter: Exclude any other duplicates

- Math Formula: Calculate a “Consensus ranking”, this is the sum of all four ranking scores round (“ASP rank” + “GOLD rank” + “Chemscore rank” + “Chem PLP rank”) and appends as a new column.
- Math Formula: Calculate if the molecules have a better “Consensus rank” than that generated for pik1 and appends a new column.
- Numeric Row Splitter: Remove any ligands with a “Consensus rank” 4, below that of pik1.
- Numeric Row Splitter: Remove any ligands with a “Solubility” (logS) less than -4.
- Pareto Ranking: List of all columns which may be used for ranking. Only numeric columns are supported currently.
- Sorter: Sort the “Pareto ranking” into ascending order.
- SD Writer: Create an output file containing all the information calculated in an SDF format (See Chapter 2.1 Figure 2.6).

6.1.5 Second Generation Inhibitors

6.1.5.1 Design Protocol 7

Second generation inhibitor design was carried out using Inhibitor 2 as a template as it scored most highly using 6.1.1 - 6.1.4, optimisation was conducted in concordance with the docking pose of the Inhibitor 2 NCS1 complex determined under 6.1.3 Molecular Docking Methods Protocols 3 and 4. Solubilising groups were then selected with the aim to improving the ADMET parameters of the new inhibitors and hence 55 structurally diverse variants of Inhibitor 2 were designed and modelled using Spartan 08 V1.2.0.[225]

6.1.5.2 Molecular Docking Methods

The analogues were docked using GOLD Suite v5.2 following 6.1.3 Molecular Docking Methods protocols 3 and 4,[99] with an amendment to the protein file used, replacing the frq1 pik1 complex (pdb 2JU0)[34] with the NCS1/ D2R peptide complex (pdb code 5AER).[267] The binding site was determined as 6 Å of “L_1, C:\backup\Helix 1 dopamine.pdb” (one dopamine c terminal helix).

6.1.5.3 ADMET Properties Protocol 8. [125]

ADMET properties of the analogues were calculated in Pipeline Pilot Student Edition version 8.5 using the following components in a sequential protocol;

- “Chemistry Readers”; “SD Reader” reads in the SD file from the Goldscore docking results calculated in 5.1.5.2 Molecular Docking Methods.[99]

- The ADMET properties were calculated using a number of components found under “Calculators”; including “Solubility”, “Surface Area and Volume”, “Number of Hydrogen Acceptors and Donors”, “A LogP, “Molecular Weight” and “Molecular Properties” specifically selected to calculate the number of rotatable bonds.
- “Chemistry Writers”; “SD Writer” and “Excel Report Writer”. Creates two output formats, one an SD file and the other lists all the information calculated under 6.1.5.2 and 6.1.5.3 in an xls file spreadsheet that can used under 6.1.5.4 “Ranking and Selection”.

6.1.5.4 Ranking and Selection

The scoring output spreadsheet (xls) was analysed using Microsoft office Excel 2007, with the ligand efficiency of each analogue calculated using the formula found within the 6.4.2 Balanced Selection Protocol 6. The analogues were then ranked in order of their “Molecular solubility” and the top 12 molecules with best predicted solubility were selected for synthesis.

6.1.6 Fragment Library

6.1.6.1 Construction

The small molecule library of 1137 compounds was initially stored in a Microsoft Office Excel® 2007(Redmond, Washington: Microsoft) format. For the purpose of computational modelling, this file was converted into a SDF format with each individual molecule in three dimensional energy minimised form. This was carried out using Pilot Student Edition version 8.5, 6.1.6.1.1 Protocol 9.[125] The SDF file was then modelled in Spartan ’08 V1.2.0 Protocol 10.[225]

6.1.6.1.1 Protocol 9[125]

- “File Reader”; “Delimited text reader”. Reads in excel files also known as delimited text files.
- “Chemistry”; “Data Access and Manipulation”; “Converters”; “Molecule from text” (simplified molecular-input line-entry system “SMILES” format). Converts the chemical name of a compound into its molecular structure specifically in a SMILES data format.
- “Chemistry”; “Data Access and Manipulation”; “Manipulators”; “Add hydrogen’s”.

- “Chemistry”; “Data Access and Manipulation”; “Manipulators”; “3-D Methods”; “3-D co-ordinates”. Creates 3-D conformer of each compound from the 2-D SMILES file.
- “Chemistry”; “Data Access and Manipulation”; “Manipulators”; “3-D Methods”; “Minimise molecule”. Creates an energy minimised conformation of each compound.
- “Chemistry Writers”; “SDF writer”. Exports files in SD format.

6.1.6.1.2 Protocol 10 [225]

- File Open SDF file.
- Setup; Calculations; Equilibrium Geometry;-Molecular mechanics. This carries out an energy minimisation optimisation.[322][225]
- Submit.
- File; save as in “SDF” and “Spartan formats”.

6.1.6.2 Molecular Docking Methods[99]

The 1137 small molecules from 6.1.6.1 were docked and rescored using GOLD Suite v5.2 following protocol 6.1.3 Molecular Docking Methods protocols 3 and 4, with an amendment to the protein file used replacing the frq1 pik1 complex (pdb 2JU0) with NCS1 D2R peptide (pdb code 5AER). [99][34][267] The binding site was determined as 6 Å of “L_1, C:\backup\Helix 1 dopamine.pdb” (one dopamine c terminal helix). The 100 docking poses for each compound were then separated using Pipeline Pilot Student Edition version 8.5 using 5.1.6.2.1 Protocol 11.[125]

6.1.6.2.1 Protocol 11

- SD Reader- Reads in the SD file.
- Custom manipulator- manipulates the data.
- “Group data by tag”-All 100 docking poses for each compound are grouped using “mol name”.
- “Sub protocol 1”, individual mol sdf files generated, Pwd: (c) \programme files\Accelrys\PPs\public\users\jh0u700c. Used to extract the 100 docking poses for each compound as one file.

6.1.6.3 Selection Methods

The individual Goldscore files were grouped for each molecule and an average Goldscore was calculated, the values were then merged with the compound library file, conducted using Pipeline Pilot Student Edition version 8.5, 6.1.6.3.1 Protocol 12.

The second step in the selection process was the calculation of the ADMET parameters the protocol followed was that of 6.1.5.3.[184] A final filter applied known as “the rule of three”,[323] those compounds with a molecular weight ≤300, AlogP ≤ 3 hydrogen bond donors and hydrogen bond acceptors ≤ 3 respectively were carried forward, those that did not fit the criteria were discarded. This process was carried out using Pipeline Pilot Student Edition version 8.5 following 6.1.6.3.2 Protocol 13.[125] Protocols 12 and 13 were then repeated for the rescoring functions ASP, Chemscore and ChemPLP.[99]

The four separate files generated from these processes were combined and the compounds ranked using Knime 2.5.1, 6.1.6.3.2 Protocol 14.[124] Finally, the SDF file generated from Protocol 14 was then recombined with the original spreadsheet.

6.1.6.3.1 Protocol 12

- “SD Reader”- Read in SD file from 6.1.6.2.1
- “Custom manipulator”- Find “molname”
- “Group data” by tag- Group data according to the “molname” column
- “Calculate average nodefine”- *Expression: #Root=DataRoot(); #SumGoldscore=(); For # I in 1… NodeNumChildren (#Root)loop; #child:= nodeithchild (#Root, #i); #SumGoldscore+ = Nodeproperty (#child,’Gold.Goldscore.Fitness’); End loop; AvgGoldscore:=SumGoldscore/NodeNumChildren(#Root)*- Calculate average scoring function such as “Goldscore”, creates column AvgGoldscore.
- “Keep properties”- molname, AvgGoldscore - Filter unnecessary data.
- “Remove group data” - Remove duplicates.
- “Rename property”- Rename “molname” to “Name”.
- “SD Reader”- Import 6.1.6.1 SD library file
    - Merge data- merge both sets of data using “Name”.
- “Output”- “SD writer” and “HTML Viewer”- Export results in two file formats SD and an online molecular viewer.

6.1.6.3.2 Protocol 13

- “SD Reader”.
- “Property value”: “threshold filter 1”- Select for compounds with a molecular solubility ≥3.
- “Sort data”.

- "Property value": "threshold filter 2" - Select for compounds with ALogP ≤ 3.
- "Sort data".
- "Property value": "threshold filter 3" - Select for compounds with molecular weight ≤300.
- "Sort data".
- "Property value": "threshold filter 4" - Select for compounds with number of hydrogen bond acceptors ≤ 3.
- "Sort data".
- "Property value": "threshold filter 5" - Select for compounds with number of hydrogen bond donor's ≤ 3.
- "Sort data".
- "Remove hydrogens".
- "2-D co-ordinates"- Generate 2-D molecular structure.
- "Output options": "Excel structure viewer", "SD writer" and "HTML molecular viewer".

6.1.6.3.3 Protocol 14

- "SD Reader"- Read in the SD file from 6.1.6.3.2 Protocol 13.
- "Row filter"- Select the column to test: "Average ASP".
- "Math formula"- Calculate the "ligand efficiency" (average score/ molecular weight).
- "Joiner"- Join the output from two scoring functions.
- "Column filter"- Delete duplicates.
- "Joiner"- Collate all four columns into one single output.
- "Math formula"- Calculate the "Better than D2" score using the following equation.

$$Y = \frac{\varepsilon G}{\varepsilon G1} + \frac{\varepsilon A}{\varepsilon A1} + \frac{\varepsilon C}{\varepsilon C1} + \frac{\varepsilon P}{\varepsilon P1}$$

Y- Better than D2 score

εG, εA, εC, εP- Compound "Ligand efficiencies" from "Goldscore", "ASP", "Chemscore" and" Chem PLP".

εG1, εA1, εC1, εP1- D2 Ligand efficiencies from "Goldscore", "ASP", "Chemscore" and "ChemPLP" respectively.

- "Numeric row splitter"- "BetterthanD2" score≤ 4.
- "Sorter"- Sort the "BetterthanD2" score in descending order.

- SD writer.

6.1.6.3.4 Clustering and data analysis

With the compound library now listed in descending order with respects to the "BetterthanD2" score calculated previously in 6.1.6.3.3 Protocol 14 it was deemed necessary to cluster the compounds. This was carried out in Pipeline Pilot Student Edition version 8.5 using a number of different protocols to compare and contrast the data.[125]

6.1.6.3.4.1 Protocol 15

- "SD reader"- Read in SD file from 6.1.6.3.3 Protocol 14.
- 1) "Cluster molecules"- Cluster the compounds with an "average number per cluster": "6 cluster centres".
    - "Histogram numerical".
    - "HTML report viewer".
    - "Pdf report writer".
- 2) "Histogram numerical"- data series 1; property "Better than D2".
    - "HTML report Viewer".
    - "Pdf report viewer".

6.1.6.3.4.2 Protocol 16

- "SD reader"- Read in SDF file from 5.1.6.3.3 Protocol 14, then split into branches 1 and 2.
- 1) "Rename property"- "BetterthanD2".
    - 3) "Cluster molecules"- Cluster into an average number of 6 molecules per cluster, select "cluster centres and members".
        - "Sort data"- Sort the cluster into ascending numbers, with the "BetterthanD2" into descending numbers.
        - "Group data by tag"- The data is grouped by each cluster.
        - "Find maximum value for each cluster"- Calculate the maximum ligand efficiency value for each cluster.
        - "Histogram numerical"- Generate a histogram with "maximum ligand efficiency"; bin size 0.5; for the ligand efficiency per cluster.
    - 4) "Cluster molecules"- Cluster with an average number of 6, select "cluster centres", and generate the mean of each cluster by selecting "mean centres and scale".

- "Sort data"- Sort the cluster into ascending numbers, with the "BetterthanD2" into descending numbers.
- "Histogram numerical"- Create a Histogram of the"BetterthanD2"; bin size 0.5; along with the ligand efficiency all molecules.

- 2) "Histogram numerical"- Create a histogram with no clustering of "BetterthanD2"; bin size 0.5; ligand efficiency all molecules
- Joining branches 1 and 2.
- 5) Tile horizontal.
- 6) Output options: HTML report viewer and pdf writer.

6.1.6.3.4.3 Protocol 17

- 1) "SD reader"- Read in SDF file from 6.1.6.3.3 Protocol 14, then split into branches 1 and 2.

1. 2) "Top N filter"- Keep top 50 molecules.
2. 3) "Sort data"- Sort the "BetterthanD2" column in descending order.
3. 4) "Cluster molecules"- Create a larger cluster with an average number per cluster of 12, select "cluster centres and members".
    - i) "Count and index data".
    - ii) "Group data by tag"- Group using cluster.
    - iii) "Select molecules".
        - "Ungroup data".
        - "Sort data"- "BetterthanD2" descending.
        - "Top N filter"- "BetterthanD2" true max number of records.
    - iv) "2-D co-ordinates"- Generate the 2-D co-ordinates of each molecule.
4. Join the "BetterthanD2" scores for the molecules from each cluster by joining the outputs from 3) and 4).
5. "Output options"- "SD writer".

6.1.6.3.4.4 Protocol 18

6. 1) "SD reader"- Read in SDF file from 6.1.6.3.5 Protocol 17 containing 102 molecules.
7. 2) "Top N filter"- Keeps top 15 molecules.
8. 3) "Sort data"- Sort the remaining data with respects to "BetterthanD2" in descending order.

9. 4) “Cluster molecules”- Cluster the remaining 87 with an average number per cluster of 6, select “cluster centres and members”.
    - ii) Group data by tag- group using cluster.
    - iii) Select molecules.
        - Ungroup data.
        - “Sort data”- Sort by “BetterthanD2” in descending order.
        - “Top N filter”- “BetterthanD2” true max number of records.
    - iv) 2-D co-ordinates.
10. Join the “BetterthanD2” scores for the molecules from each cluster by joining the outputs from 3) and 4).

6.1.6.3.4.5 Protocol 19

- 1) “SD reader”- Read in SDF file from 6.1.6.3.6 Protocol 18
    - Tag data.
- 2) “Delimited text reader”- Read in original excel spreadsheet from 6.1.6.1
    - Rename property- changes “molecular name” to “name”.
- “Merge data”- Combines data from 1 and 2 using ‘name’ as the recognition point.
- “Property defined filter”.
- “Output”: “HTML molecular table viewer”, “SD writer”, “Delimited text writer”.

6.2 Chemical Procedures

6.2.1 Analysis Techniques and Reagents

The following general techniques for organic synthesis and analysis in this experimental chapter were employed;

6.2.1.1 Thin layer chromatography (TLC*)*

Analytical thin layer chromatography was performed on silica gel 60 coated aluminium TLC plates. Once developed the chromatographs were detected and analysed using a UV lamp (254 nm) and stained using an appropriate dip, developed using a heat gun.

6.2.1.2 Flash column chromatography

The required amount of silica (particle size 40 – 63 µm, supplied by Aldrich) was made into slurry with the appropriate volume of the desired eluent system; it was then applied to a column over a thin base layer of sand. The crude product was applied to the column, either dissolved in the minimum volume of the eluent system or pre absorbed onto a small amount of silica. The column was then eluted with the eluent and fractions were collected and analysed using TLC.

6.2.1.3 Nuclear Magnetic Resonance spectroscopy (NMR)

^{1}H and ^{13}C NMR spectra were both recorded on a Bruker AMX-400 Spectrometer in deuterated solvents as indicated within the experimental data, ^{1}H spectra were recorded at operating frequency of 400 MHz and ^{13}C at 100 MHz. The chemical shifts are reported in ppm and the coupling constants (*J*) in hertz (Hz).

6.2.1.4 Mass Spectrometry

Mass Spectrometry was performed using an in house service within the Chemistry Department of the University of Liverpool and was conducted on a VG analytical 7070E machine and a Frisions TRIO spectrometers using electronic ionisations (EI), chemical ionisations (CI) and electro spray (ES).

6.2.1.5 Solvents

Analytical grade solvents were prepared for reactions involving the use of semi aqueous media, anhydrous or purified solvents were obtained in the following manner:

Tetrahydrofuran (THF)- Dried from a sodium and benzophenone still under nitrogen for 2-3 hours until purple.

Diethyl ether- Dried from a sodium and benzophenone still under nitrogen for 2-3 hours until purple.

Toluene- Sureseal™ anhydrous solvent purchased from Aldrich, used without modification.

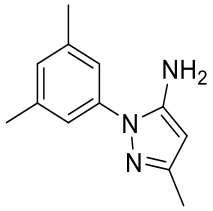
N,N-Dimethylformamide (DMF)- Sureseal™ anhydrous solvent purchased from Aldrich or distilled from calcium hydride before use.

Dimethylacetamide (DMA)- Sureseal™ anhydrous solvent purchased from Aldrich, used without modification.

Methanol (MeOH)- Sureseal™ anhydrous solvent purchased from Aldrich or dried from a magnesium with iodine still under nitrogen for 2-3 hours.

Ethanol (EtOH)- Dried from a magnesium with iodine still under nitrogen for 2-3 hours.

6.2.2 1-(3,5-dimethylphenyl)-3-methyl-1H-pyrazole-5-amine[245]

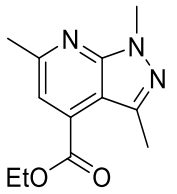


2.11

A solution of 3-aminocrotonitrile (1.25g, 15.2 mmol) in HCl (1M, 50 mL) and 3, 5-dimethyl-phenylhydrazine (2.5g, 15.2 mmol) were refluxed at 115 °C for four hours. The solution was then cooled to 25 °C and made basic by the drop wise addition of NaOH (6M). The reaction was extracted with DCM (3 × 50 mL), the organic fractions were combined and dried with anhydrous $MgSO_4$. The solvent was removed *in vacuo* to afford an off white solid without further purification (2.00 g, 9.94 mmol, 80.3%).

R_f = 0.44 (1:9 MeOH: DCM). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.12 (2H, s, Ar CH), 6.95 (1H, s, Ar CH), 5.42 (1H, s, CH), 3.79 (2H, br s, NH_2), 2.34 (6H, s, 2×CH_3), 2.22 (3H, s, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 149.5, 145.6, 139.7, 138.8, 129.2, 121.9, 90.8, 21.6, 14.3. **NRMS** Calculated for $C_{12}H_{15}N_3$ + H^+ [M + H^+]: 202.2 Da. Found: [M + H^+] 202.2 Da. **Elemental analysis** calculated (%) for $C_{12}H_{15}N_3$: C 71.61, H 7.51, N 20.88. Found: C 71.23, H 7.58, N 20.71.

6.2.3 1,3,6-trimethyl-1H-pyrazolo-3,4-pyridine-ethylester[246]

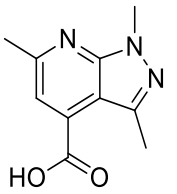


2.14

1,3-Dimethyl-1H-pyrazole-5-amine (2.0 g, 17.9 mmol) and ethyl acetopyruvate (2.53 mL, 17.9 mmol) in AcOH (90 mL) were refluxed for four hours. The reaction was allowed to cool to 25 ˚C and the AcOH removed *in vacuo.* To afford the crude product, a dark brown oil which was extracted into DCM (20 mL) and washed once with water (20 mL) and once with sodium bicarbonate (20 mL). The organic layer was dried over anhydrous $MgSO_4$ and the solvent removed in *vacuo* yielding a brown solid as a crude product without further purification (3.5 g, 15.01 mmol, 83.4%).

R_f = 0.27 (1:1 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.40 (1H, s, Ar CH), 4.40 (2H, q, *J=7.2 Hz*, CH_2), 4.04 (3H, s, *N*-CH_3), 2.67 (3H, s, CH_3), 2.65 (3H, s, CH_3), 1.40 (3H, t, *J=7.2 Hz*, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 169.5, 158.6, 152.7, 140.6, 133.6, 117.7, 109.5, 62.2, 33.9, 25.1, 16.5, 14.6. **NRMS** Calculated for $C_{12}H_{15}N_3O_2$ + H^+ [M + H^+] 234.2 Da. Found: [M + H^+] 234.2 Da. **Elemental analysis** calculated (%) for $C_{12}H_{15}N_3O_2$: C 61.79, H 6.48, N 18.01. Found: C 61.65, H 6.53, N 17.92.

6.2.4 1,3,6-trimethyl-1H-pyrazolo-3,4-pyridine-carboxylic acid[246]

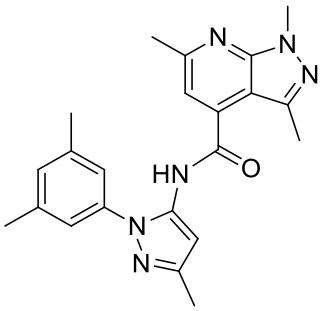


2.15

To a stirring solution of **2.14** (3.5 g, 14.4 mmol) in *i*-PrOH (100 mL), was added KOH (1.68 g, 29.8 mmol) and the reaction refluxed for 5 hours. The solution was allowed to cool to 25 ˚C, acidified by the drop wise addition of HCl (2 M), precipitating a white solid. The solution was then partitioned between water (20 mL) and DCM (20 mL) and the pure product filtered and dried under Kugelrohr conditions yielding a fine white powder without further purification (2.21 g, 10.7 mmol, 72%).

R_f =0.08 (1:9 MeOH: DCM) **^{1}H NMR** (400 MHz, MeOD): δ 7.51 (1H, s, CH), 4.05 (3H, s, *N*-CH_3), 2.72 (3H, s, CH_3), 2.66 (3H, s, CH_3). **^{13}C NMR** (100 MHz, MeOD): δ 167.2, 158.3, 151.9, 139.5, 134.4, 117.1, 108.8, 33.6, 24.5, 15.8. **NRMS** Calculated for $C_{10}H_{11}N_3O_2 + H^+$ [M + H^+]: 206.2 Da. Found: [M + H^+] 206.2 Da. **Elemental analysis** calculated (%) for $C_{10}H_{11}N_3O_2$: C 58.53, H 5.40, N 20.48. Found: C 58.52, H 5.41, N 20.38.

6.2.5 Inhibitor 1



Inhibitor 1

6.2.5.1 Procedure 1[246]

Acid **2.15** (200 mg, 1.0 mmol), amine **2.11** (210 mg, 1.1 mmol), *N,N*-diisopropylethylamine (DIPEA) (0.505 mL, 2.9 mmol) and 1-methyl-2chloropyridinium iodide (370 mg, 1.5 mmol) were refluxed in anhydrous MeCN (3 mL) for 4 hours. The solvent removed *in vacuo* and the crude product analysed using ^{1}H NMR, without further purification (88.1 mg, 0.2 mmol, 23%).

Rotamers

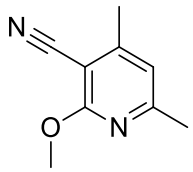
R_f =0.22 (1:1 EtOAc:Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): *Major peaks* δ 8.10 (1H, s, NH), 7.05 (1H, s, CH), 6.91 (1H, s, CH), 6.81 (1H, s, CH), 6.61 (1H, s, CH), 4.0 (3H, s, CH_3), 2.61 (3H, s, CH_3), 2.44 (3H, s, CH_3), 2.32 (3H, s, CH_3), 2.31 (6H, s, 2× CH_3). *Minor peaks* δ 8.40 (1H, s, NH), 6.87 (1H, s, CH), 6.81 (1H, s, CH), 6.35 (1H, s, CH), 6.12 (1H, s, CH), 2.49 (3H, s, CH_3), 2.25 (3H, s, CH_3), 2.20 (3H, s, CH_3), 2.17 (6H, s, 2× CH_3).

6.2.5.2 Procedure 2

To a stirring solution of acid **2.15** (200 mg, 1.0 mmol) in anhydrous THF (10 mL), DCC (310.6 mg, 1.5 mmol) and HOBt (197.5 mg, 1.5 mmol) were added and allowed to react at 25 ˚C for 30 minutes. Amine **2.11** (215.6 mg, 1.1 mmol) was added and the reaction was left stirring at 25 ˚C for 24 hours. The reaction was followed by TLC and after 24 hours it was deemed necessary to heat the reaction to 50 °C for a further 24 hours. The crude product was partitioned between water (1× 50 mL) and DCM (3×50 mL) and purified using column chromatography in a solvent system of concentration gradient of EtOAc in Hexane. Yielding the product which appeared as an off white powder (109 mg, 0.3 mmol, 28%).

R_f= 0.22 (1:1 EtOAc: Hex)**.** **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.90 (1H, br s, NH), 7.10 (1H, s, CH), 7.02 (1H, s, CH), 6.95 (1H, s, CH), 6.68 (1H, s, CH), 4.04 (3H, s, N-CH_3), 2.60 (3H, s, CH_3), 2.50 (3H, s, CH_3), 2.40 (3H, s, CH_3), 2.30 (6H, s, 2× CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 169.5, 163.6, 152.2, 149.8, 140.2, 140.1, 137.9, 137.4, 135.4, 130.5, 122.6, 114.2, 109.0, 99.7, 33.9, 25.1, 21.6, 14.8, 14.3. **IR** $\tilde{\nu}$ = 3200 (m; *v* (N-H)), 1662 (s; *v* (C=O)). **HRMS** Calculated for $C_{22}H_{24}N_6O$ + Na^+ [M + Na^+]: 411.1909 Da. Found [M + Na^+]: 411.1912 Da (+0.7 ppm). **Elemental analysis** calculated (%) for $C_{22}H_{24}N_6O$: C 68.0, H 6.2, N 21.6. Found: C 67.3, H 6.1, N 20.5.

6.2.6 2-methoxy-4,6-dimethylnicotino-3-nitrile[251]

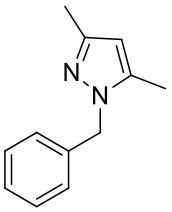


2.21

2-Hydroxy-4,6-dimethylnicotino-3-nitrile (1.0 g, 6.7 mmol) was dissolved in DCM (52 mL). To this solution silver carbonate (2.61 g, 9.5 mmol) and methyl iodide (4.41 mL, 70.8 mmol) were added with protection from light. The reaction was stirred at 25 ˚C for 24 hours, then the solvent was removed *in vacuo* to afford the crude product which was purified using column chromatography in a solvent system of EtOAc and Hexane. The pure product appeared as a fine white powder (767 mg, 4.7 mmol, 70.1%).

R_f= 0.83 (7:3 EtOAc: Hex). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 6.68 (1H, s, CH), 4.01 (3H, s, OCH_3), 2.45 (3H, s, CH_3), 2.44 (3H, s, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 164.6, 161.1, 154.6, 117.8, 115.5, 94.2, 54.6, 24.9, 20.4. **NRMS** Calculated for $C_9H_{10}N_2O$ $+H^+$ [M + H^+]: 163.2 Da. Found [M + H^+]:163.1 Da. **Elemental analysis** calculated (%) for $C_9H_{10}N_2O$: C 66.65, H 6.21, N 17.27. Found: C 66.42, H 6.24, N 17.05.

6.2.7 *N*-Benzyl-3,5-dimethylpyrazole[324]

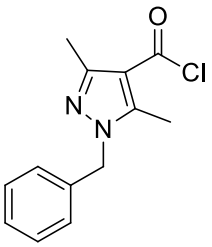


2.23

3,5-Dimethylpyrazole (1.0 g, 10.5 mmol) was added to a solution of KOH (0.88 g, 15.6 mmol) in DMSO (14.5 mL). The mixture was stirred at 25 ˚C and then heated at 80 °C for one hour, before being cooled to 25 ˚C and benzyl chloride (1.2 mL, 10.4 mmol) was added. The reaction mixture was stirred at 25 ˚C for a further two hours. The mixture was poured into water (50 mL) and extracted into DCM (4 × 50 mL). The organic layers were combined, washed with water (4 × 50 mL), saturated NaCl (1 × 50 mL), dried over anhydrous sodium sulphate and the solvent was removed *in vacuo*. The crude product was purified using column chromatography in a solvent system of EtOAc and Hexane to yield the product which appeared as a yellow oil (1.35 g, 7.3 mmol, 70%).

R_f = 0.22 (5:95 EtOAc: Hexane). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.30-7.00 (5H, m, Ph), 5.80 (1H, s, CH), 5.20 (2H, s, CH_2), 2.20 (3H, s, CH_3), 2.10 (3H, s, CH_3). **^{13}CNMR** (100 MHz, $CDCl_3$): δ 147.9, 139.6, 137.8, 129.9, 127.8, 126.9, 105.9, 53.0, 13.9, 11.5. **IR** $\tilde{\nu}$ = 1554 (m; *v* (C=C)), 1455 (m; *v* (C=C)). **NRMS** Calculated for $C_{12}H_{14}N_2 + H^+$ [M + H^+]: 187.2 Da. Found [M + H^+]: 187.2 Da.

6.2.8 *N*-Benzyl-3,5-dimethyl-1-H-pyrazole-4-carbonylchloride[253]

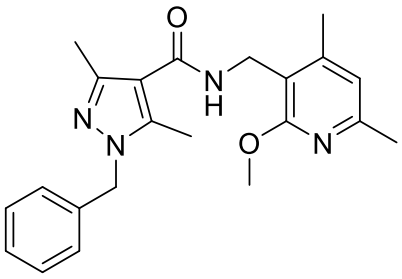


2.24

Acid **2.23** (1.4 g, 7.3 mmol) and oxalyl chloride (2.5 mL, 29 mmol) were refluxed for 3 hours. The excess oxalyl chloride was removed *in vacuo* yielding the crude product which appeared as a thick orange oil and was carried forward into the next step without purification and full characterisation.

IR $\tilde{\nu}$ =: 1735 (s; *v* (C=O)), 1616 (m; *v* (C=C)), 1427 (m; *v* (C=C)), 701 (m; *v* (C-Cl)).

6.2.9 Inhibitor 2[325]

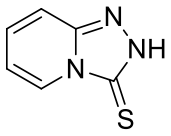


Inhibitor 2

Nitrile **2.21** (0.3 g, 1.8 mmol) dissolved in the minimum amount of anhydrous Et_2O (3 mL) was added to a slurry of lithium aluminium hydride (0.14 g, 3.7 mmol) in anhydrous Et_2O (3 mL). The reaction was stirred at 25˚C for two hours and then heated for two hours at reflux. The excess hydride was decomposed by the careful addition of the minimum amount of 20% NaOH (2 mL). Crude acid chloride **2.24** was then added slowly (1.6 g, 5.6 mmol) along with pyridine (2 mL) and the reaction was left stirring at 25 ˚C for 16 hours. Upon completion the mixture was basified with the addition of aqueous $NaHCO_3$, filtered and washed with hot toluene (2 × 10 mL) and DCM (2 × 10 mL). The filtrate was extracted into DCM (3× 10 mL) and the combined organic fractions were dried over anhydrous $MgSO_4$. The solvent was removed *in vacuo* and purified using column chromatography, in a solvent system of EtOAc and Hexane. Yielding the product which appeared as a fine white powder (225 mg, 0.6 mmol, 32%).

R_f = 0.20 (50:50 EtOAc: Hex)) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.30-7.03 (5H, m, Ph), 6.50 (1H, s, CH), 5.20 (2H, s, CH_2), 3.90 (3H, s, O-CH_3), 2.42 (3H, s, CH_3), 3.40 (3H, s, CH_3), 2.37 (6H, s, 2 × CH_3) **^{13}C NMR** (100 MHz, $CDCl_3$): δ 164.5, 162.3, 154.7, 148.2, 146.2, 142.4, 136.6, 129.2, 128.1, 127.0, 119.1, 116.4, 114.3, 53.7, 53.1, 34.8, 30.1, 24.2, 19.4, 14.4, 11.3. **HRMS** Calculated for $C_{22}H_{26}N_4O_2$ + Na^+ [M + Na^+]: 401.1953 Da. Found [M + Na^+]: 401.1964 Da (-2.9ppm). **Elemental analysis** calculated (%) for $C_{22}H_{26}N_4O_2$: C 69.8, H 7.06, N 14.8. Found: C 69.3, H 7.1, N 14.4.

6.2.10 (1,2,4)-triazolo (4,3-a) pyridine-3 (2H)-thione[326]

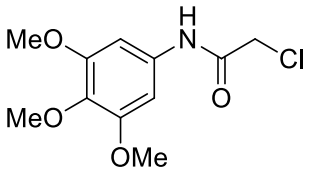


2.30

2-Hydrazinopyridine (1.0 g, 9.16 mmol) was added to a solution of carbon disulfide (2.4 mL, 40.3 mmol) in $CHCl_3$. The reaction was refluxed for 20 hours, after which time the product had crystallised out of solution. The crystals were filtered and dried under vacuum yielding the product which appeared as fine off white crystals with no further purification (1.1 g, 7.0 mmol, 76.4%).

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.35 (1H, dt, *J=7.0, 1.0 Hz*, CH), 7.49 (1H, dt *J=9.0, 1.1 Hz*, CH_4), 7.36 (1H, ddd, *J=9.0, 6.0, 1.2 Hz*, CH_3), 6.87 (1H, ddd *J=7.5, 6.0, 1.0 Hz* ,CH_2). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 159.5, 146.5, 131.4, 125.4, 115.9, 114.2**. NRMS** Calculated for $C_6H_5N_3S$ + H^+ [M + H^+]: 152.2 Da. Found [M + H^+]: 152.2 Da. **Elemental analysis** calculated (%) for $C_6H_5N_3S$: C 47.66; H 3.33; N 27.79. Found: C 47.76, H 3.31, N 27.98.

6.2.11 2-chloro-*N*-(3,4,5- trimethoxyphenyl)acetamide[327]

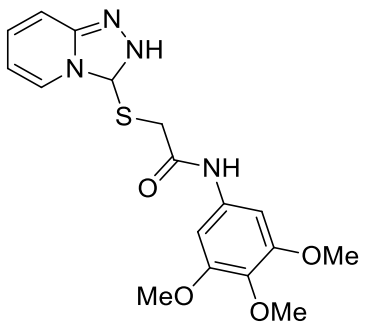


2.33

Chloroacetyl chloride (0.6 mL, 8.2 mmol) and NEt_3 (1 mL) were added to a stirring solution of 3, 4, 5-trimethoxy aniline (1.0 g, 458 mmol) in anhydrous toluene (20 mL). The reaction was refluxed for 15 hours forming a dark brown solid precipitate. All solvents were removed *in vacuo*, the crude mixture was re-suspended in DCM and washed with toluene (3 × 10 mL). The DCM was then removed *in vacuo* and the crude product purified using column chromatography, yielding the product which appeared as an off white powder (1.2 g, 4.6 mmol, 83.9%).

R_f =0.50 (80:20 EtOAc: Hex). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.17 (1H, br s, NH), 6.840 (2H, s, Ar CH), 4.19 (2H, s, CH_2), 3.86 (6H, s, 2×OMe), 3.83 (3H, s, OMe). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 164.6, 153.6, 135.4, 133.5, 98.3, 61.3, 56.4, 43.4. **HRMS** Calculated for $C_{11}H_{14}ClNO_4$ + Na^+ [M + Na^+]: 282.0509 Da. Found [M + Na^+] 282.0503 Da (-2.1 ppm). **Elemental analysis** calculated (%) for $C_{11}H_{14}ClNO_4$: C 50.88, H 5.43, N 5.39. Found: C 51.07; H 5.47; N 5.31.

6.2.12 Inhibitor 3

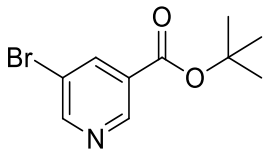


Inhibitor 3

N,N-diisopropylethylamine (0.54 mL, 3.1 mmol) was added to a stirring solution of thione **2.30** (815 mg, 5.4 mmol) and chloramide **2.31** (200 mg, 0.7 mmol) in anhydrous DMF (10 mL) and the reaction heated to 120 °C for one hour. The product was purified using flash column chromatography in a solvent system of EtOAc and Hexane. Yielding the product with appeared as an off white solid (0.189 g, 0.5 mmol, 65%).

R_f = 0.35 (50:50 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 10.20 (1H, br s, NH), 8.10 (1H, d, *J= 7.0 Hz,* CH), 7.80 (1H, d, *J= 9.3 Hz* CH), 7.36 (1H, ddd, *J= 9.3, 6.6, 1.0 Hz*, CH), 6.97 (1H, td, *J = 6.9, 0.8 Hz*, CH), 6.90 (2H, s, 2×CH), 4.10 (2H, s, CH_2), 3.80 (6H, s, 2×OMe), 3.70 (3H, s, OMe). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.5, 153.6, 151.2, 142.1, 135.1, 143.6, 128.9, 123.5, 116.6, 115.2, 97.9, 61.3, 56.5, 38.5. **HRMS** Calculated for $C_{11}H_{14}ClNO_4$ + Na^+ [M + Na^+]: 397.0946 Da. Found [M+ Na]$^+$ 397.0960 Da (+3.4 ppm). **Elemental analysis** calculated (%) for $C_{11}H_{14}ClNO_4$: C 54.5; H 4.9; N 14.9. Found C 54.7; H 4.9; N 14.2.

6.2.13 *tert*-Butyl-5-bromonicotinate[328]

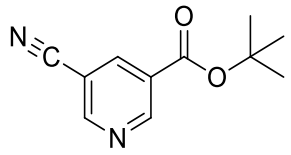


3.15

1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (8.43 g, 54.3 mmol) was added to a stirring solution of 2,5-bromonicotinic acid (10 g, 49.4 mmol) in $CHCl_3$ (100 mL), dimethyl amino pyridine (DMAP) (6.6 g, 54.3 mmol) and *t*-butanol (18.6 mL, 197.6 mL). The reaction was stirred at 25 ˚C for 72 hours and purified using flash column chromatography in a solvent system of EtOAc and Hexane. Yielding the product which appeared as a white solid (7.3 g, 28.7 mmol, 57%).

R_f = 0.87 (50:50 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 9.06 (1H, s, CH), 8.80 (1H, s, CH), 8.35 (1H, s, CH), 1.60 (9H, s, 3×CH_3) **^{13}C NMR** (100 MHz, $CDCl_3$): δ 163.05, 154.01, 148.84, 139.39, 129.05, 120.47, 28.10. **HRMS** Calculated for $C_{10}H_{12}BrNO_2$ + H^+ [M + H^+]: 258.0124 Da. Found [M + H^+]: 258.0126 Da (-0.65 ppm). **Elemental analysis** calculated (%) for $C_{10}H_{12}BrNO_2$: C 46.53, H 4.69, N, 5.43. Found: C 46.91, H 4.69, N 5.24.

6.2.14 *tert*-Butyl-5-cyanonicotinate[329]

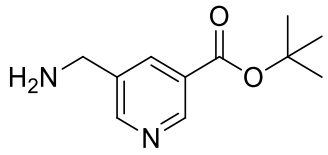


3.16

To bromopyridine **3.15** (500 mg, 1.9 mmol) in anhydrous dimethylacetamide (12.15 mL) $Pd(dba)_2$ (88.5 mg, 0.1 mmol), dppf (53.6 mg, 0.1 mmol), zinc cyanide (250.3 mg, 2.1 mmol) and zinc (139.3 mg, 2.1 mmol) were added. The reaction was heated to 120 ˚C for 5 hours, the reaction solution was filtered through celite and extracted into EtOAc (30 mL), washed with saturated NaCl (3 x 30ml) and dried over anhydrous $Na_2SO_4$the solvent removed *in vacuo*. The product purified using flash column chromatography in a solvent system of EtOAc and Hexane yielding the product an off white solid (281.9 mg, 1.4 mmol, 71.3%).

R_f = 0.79 (1:1 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 9.25 (1H, d, *J= 1.9 Hz*, CH), 8.93 (1H, d, J= 1.9 Hz, CH), 8.43 (1H, t, J= 2.0 Hz, CH), 1.55 (9H, s, 3×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 162.23, 154.95, 153.64, 140.15, 127.90, 115.81, 109.93, 83.57, 28.03. **IR *ṽ*** = 2234 (m; *v* (C≡N)), 1707 (s; *v* (C=O)), 1597 (m; *v* (C=C)), 1443 (m; *v* (C=C)). **HRMS** Calculated for $C_{11}H_{13}N_2O_2 + H^+$ [M + H^+]: 205.0972 Da. Found [M + H^+] 205.0972 Da (-0.35 ppm). **Elemental analysis** calculated (%) for $C_{11}H_{13}N_2O_2$: C 64.69, H 5.92, N 13.72 Found C 65.41, H, 5.91, N 13.16.

6.2.15 *tert*-Butyl-5-(aminomethyl)nicotinate[287]

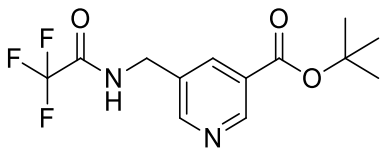


3.17

25% NH_3 solution (304 µL) and a water slurry of Raney nickel (304 µL, 4.9 mmol) was added to a solution of nitrile **3.16** (100 mg, 0.5 mmol) in EtOH (4 mL) and THF (4 mL). The reaction was degassed and under a hydrogen atmosphere stirred at 25 ˚C for 3 hours. The solution was filtered and the solvent removed *in vacuo*, the orange residue was partitioned between EtOAc (10 mL) and $NaHCO_3$ (10 mL), the organic layer was then washed with saturated NaCl and dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo*. The crude product was not purified but carried forward to the subsequent step, it appeared as an orange oil and was analysed using ^1H, ^{13}C NMR and mass spectrometry before use.

R_f = 0.53 (1:9 MeOH: DCM). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 9.01 (1H, s, CH), 8.68 (1H, s, CH), 8.19 (1H, s, CH), 3.95 (2H, s, CH_2), 2.94 (2H, br s, NH), 1.58 (9H, s, 3×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 165.37, 153.13, 150.38, 136.88, 128.62, 83.11, 44.28, 29.12. **HRMS** Calculated for $C_{11}H_{17}N_2O_2 + H^+$ [M + H^+]: 209.1285 Da. Found [M + H^+]: 209.1290 Da (-2.67 ppm).

6.2.16 *tert*-Butyl 5-((2,2,2-trifluoroacetamido)methyl)nicotinate[330]

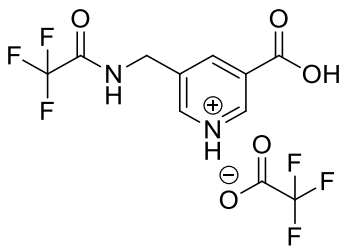


3.18

Crude **3.17** (700 mg, 3.4 mmol) was dissolved in DCM (21 mL), to this solution pyridine (1.6 mL, 20.2 mmol) and trifluoro acetic anhydride (2.4 mL, 16.8 mmol) were added. The reaction was stirred at 25 ˚C for 16 hours, diluted further with DCM and washed with water. The organic layer was dried over Na_2SO_4 and the solvent removed *in vacuo* and purified using flash column chromatography to yield the pure product which appeared as an off white crystalline solid (701.9 mg, 2.3 mmol, 68%).

R_f = 0.46 (1:1 EtOAc: Hex). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 9.03 (1H, s, CH), 8.64 (1H, s, CH), 8.15 (1H, s, CH), 7.69 (1H, br s, NH), 4.58 (2H, d, *J= 5.8 Hz*, CH_2), 1.58 (9H, s, 3×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 163.86, 125.24, 150.24, 136.56, 131.86, 127.98, 117.21, 114.35, 82.71, 41.00, 28.06. **HRMS** Calculated for $C_{13}H_{15}F_3N_2O_3$ + Na^+ [M + Na^+]: 327.0932 Da. Found [M + Na^+]: 327.0924 Da (-2.6 ppm).

6.2.17 5-((2,2,2-trifluoroacetamideo)methyl)nicotinic acid[331]

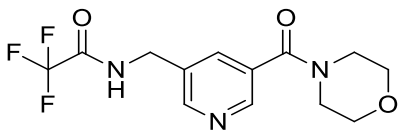


3.19

3.18 (100 mg, 0.328 mmol) was dissolved in 1.5 mL of DCM, to this stirring solution was added trifluoroacetic acid (1.5 mL, 19.6 mmol). The reaction was stirred at 25 ˚C for 3 hours, the solvent removed *in vacuo* and the residue re-suspended in water (5 mL) and extracted into EtOAc (3 × 5 mL). The organic layer was dried over Na_2SO_4, filtered and the solvent removed *in vacuo* to afford the crude product which appeared as a pale yellow solid (54.1 mg, 0.2 mmol).

R_f = 0.10 (2:8 MeOH: DCM). **^{1}H NMR** (400 MHz, MeOD): δ 9.17 (1H, s, CH), 8.87 (1H, s, CH), 8.66 (1H, s, CH), 4.67 (2H, s, CH_2). **^{13}C NMR** (100 MHz, MeOD): δ 173.10, 166.18, 150.05, 147.73, 141.84, 136.99, 130.08, 115.96, 41.34. **HRMS** Calculated for $C_9H_8F_3N_2O_3$ + Na^+ [M+ Na^+]: 249.0482 Da. Found [M+ Na^+]: 249.0482 Da (-3.52 ppm).

6.2.18 2,2,2-trifluoro-N-((5-(morpholino)pyridin-3-ylmethanone[288,332]

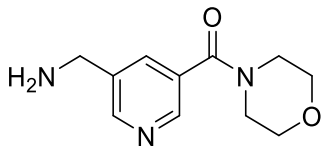


3.20

3.19 (1.03 g, 2.8 mmol) was dissolved in anhydrous DMF (11.6 mL), to this solution HATU (1.9g, 5.1 mmol) was added the resultant mixture was stirred at 25 ˚C for 30 minutes. After this time DIPEA (1.9 mL, 11.4 mmol) was added and again the reaction was left stirring at 25 ˚C for 30 minutes. Finally morpholine (0.6 mL, 7.1mmol) was added, the reaction was heated to 40 ˚C for 24 hours, cooled to 25 ˚C and stirred for a further 48 hours. The solvent was removed *in vacuo,* the residue was re-suspended in EtOAc (20 mL) and washed with $NaHCO_3$ (3 x 20 mL) and water (3 x 10 mL). The organic layer was dried over anhydrous Na_2SO_4 and the solvent removed *in vacuo*, the crude product was purified using flash column chromatography to afford the pure product which appeared as a pale orange oil (496.3 mg, 1.5 mmol, 55%).

R_f = 0.35 (1:19 MeOH: DCM). **^{1}H NMR** (400 MHz, CDCl3): δ 8.52 (1H, s, CH), 8.48 (1H, s, CH), 8.32 (1H, br s, NH), 7.62 (1H, s, CH), 4.52 (2H, d, *J=6.0 Hz*, CH_2), 3.77-3.41 (8H, m, 4×CH_2). **^{13}C NMR** (100 MHz, CDCl3): δ 167.44, 158.05-157.68, 150.47, 147.18, 134.76, 132.76, 131.30, 120.26, 111.68, 66.80, 42.84. **HRMS** Calculated for $C_{13}H_{15}F_3N_3O$ + Na^+ [M+ Na^+]: 318.106 Da. Found [M+ Na^+]: 318.1069 Da (-2.79 ppm).

6.2.19 (5-(aminomethyl)pyridin-3-yl)(morpholino)methanone[333]

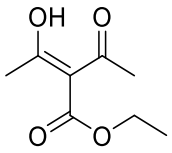


3.21

3.20 (320.4 mg, 1.4 mmol) was dissolved in 7 N methanolic ammonia (5.5 mL), the reaction was stirred at 25 ˚C for 96 hours. After this time the solvent was removed *in vacuo* and the crude residue purified using column chromatography (72.3 mg, 0.4 mmol, 31%).

R_f = 0.18 (1:19 MeOH: DCM & 0.5% NEt_3). **^{1}H NMR** (400 MHz, MeOD): δ 8.64 (1H, s, CH), 8.53 (1H, s, CH), 7.91 (1H, s, CH), 3.92 (2H, s, CH_2), 3.77-3.66 (8H, Br M, 4×CH_2). **^{13}C NMR** (100 MHz, MeOD): δ 168.02, 149.66, 145.88, 137.41, 134.63, 131.49, 66.33, 52.21, 42.04. **HRMS** Calculated for $C_{11}H_{16}N_3O_2$ + H^+ [M + H^+]: 222.1243 Da. Found [M + H^+]: 222.1240 Da (-1.1 ppm).

6.2.20 (E)-ethyl 2-acetyl-3-hydroxybut-2-enoate[290]

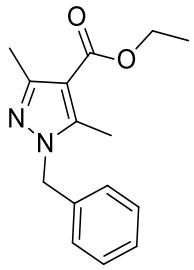


3.25

Magnesium turnings (1.9 g, 78.6 mmol), ethyl acetoacetate (10 mL, 78.5 mmol), anhydrous EtOH (15.5 mL), carbon tetrachloride (2 mL) and anhydrous toluene (197.4 mL) were combined under a nitrogen atmosphere. The reaction was left to stir at 25 ˚C for 45 minutes and then heated to reflux for one hour. Upon completion the reaction mixture was then cooled to 0 °C and acetyl chloride (5.6 mL, 79.6 mmol) was added drop wise over a period of 30 minutes to solution, after which time it was then allowed to warm to 25 ˚C and stirred for a further hour. 5% aqueous HCl (125 mL) was added to the reaction at 0 °C. The organic layer was further washed with saturated $NaHCO_3$ solution (3 × 50mL) and saturated NaCl solution (3 × 50mL), dried over $MgSO_4$ and the solvent removed *in vacuo*. This resulted in the crude product a yellow liquid which was subsequently purified using column chromatography (7.7 g, 45.12 mmol, 59%).

R_f = 0.28 (1:19 EtOAc:Hex). **^{1}H NMR** (400MHz, $CDCl_3$): δ 4.25 - 4.31 (2H, q, *J= 7.15 Hz*, CH_2), 2.38 (6H, S, 2×CH_3), 1.35 (3H, t, *J= 7.15 Hz*, CH_3). **^{13}C NMR** (400MHz, CDCl3): δ 196.57, 167.15, 108.64, 60.69, 25.89, 14.22. **NRMS** Calculated for $C_8H_{12}O_4$ + NH_3^+ [M + NH_3^+]:190.1 Da. Found [M + NH_3^+] 190.1 Da.

6.2.21 Ethyl *N*-benzyl-3,5-dimethyl-1H-pyrazole-4-carboxylate[290]

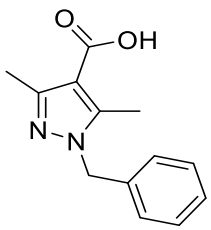


3.26

3.25 (2.2 g, 16.2 mmol) was dissolved in anhydrous EtOH (20 mL), the reaction was cooled to 0 °C and benzyl hydrazine monohydrochloride (2.8 g, 17.8 mmol) was added over 10 minutes. The reaction was warmed to 25 ˚C for 30 minutes and refluxed for 3 hours. Upon completion the solvent was removed *in vacuo* and the crude mixture, purified by flash column chromatography. (2.2 g, 8.3 mmol, 51%).

R_f = 0.77 (6:4 EtOAc:Hex). **^{1}H NMR** (400MHz, $CDCl_3$): δ 7.08-7.34 (5H, M, CH)), 5.25 (2H, S, CH_2), 4.26-4.31 (2H, q, *J=7.1 Hz*, CH_2), 2.45 (3H, S, CH_3), 2.44 (3H, S, CH_3), 1.35 (3H, t, *J=7.1 Hz*, CH_3). **^{13}C NMR** (400MHz, CDCl3): δ 164.63, 150.64, 144.20, 136.16, 128.8, 127.81, 126.66, 110.11, 59.63, 52.83, 14.40, 14.38, 11.34. **NRMS** Calculated for $C_{15}H_{18}N_2O_2$ $+H^+$ [$M+H^+$]: 259.14 Da. Found [$M+H^+$] 259.15 Da.

6.2.22 *N*-benzyl-3,5-dimethyl-1H-pyrazole-4-carboxylic acid[290]

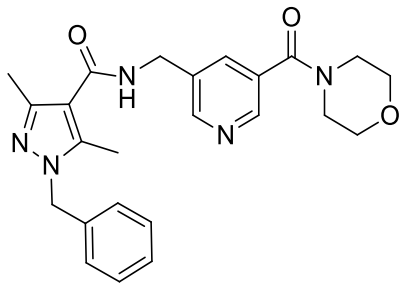


3.27

3.26 (1.3 g, 5.03 mmol) was dissolved in MeOH (17.3 mL) and 1M NaOH solution (17.3 mL, 17.3 mmol) was added. The mixture was refluxed for 1 hour and the solvent evaporated *in vacuo*. The crude product was taken up into distilled water (20 mL), the aqueous layer was then washed with Et_2O (3 × 20 mL) and cooled 0 °C. Concentrated H_2SO_4 was added drop wise, the pure carboxylic acid precipitated out and was filtered and washed further with diethyl ether (5 mL) (1.0 g, 4.3 mmol, 86%).

R_f = 0.46 (6:4 EtOAc:Hex). **^{1}HNMR** (400MHz, $CDCl_3$): δ 7.11-7.35 (5H, M, CH), 5.29 (2H, S, CH_2), 2.49 (3H, S, CH_3), 2.48 (3H, S, CH_3). **^{13}CNMR** (400MHz, $CDCl_3$): δ 169.72, 151.62, 145.35, 135.95, 128.88, 127.90, 126.72, 109.31, 52.90, 14.30, 11.43. **IR** (neat) $\tilde{\nu}$ = 2511 (s; br,*v*(O-H)), 1682 (s; *v*,(C=O)), 1577 (m; *v* (C=C)), 1533 (m; *v* (C=N)), 1458 (m; *v* (C=C)). **HRMS** Calculated for $C_{13}H_{15}N_2O_2 + H^+$ [M+H$^+$]: 231.1128 Da. Found [M+H$^+$]: 231.1.1123 Da (+2.25 Da).

6.2.23 Inhibitor 4[288,332]

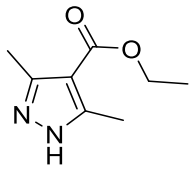


Inhibitor 4

3.27 (103.5 mg, 0.5 mmol) and HATU (346.0 mg, 0.9 mmol) were stirred in anhydrous DMF (1.3 mL) at 25 ˚C for 20 minutes, after this time potassium carbonate (K_2CO_3) (250 mg, 1.8 mmol) was added and the reaction stirred for a further 30 minutes at 25 ˚C. **3.21** (100 mg, 0.5 mmol), dissolved in anhydrous DMF (1 mL) was then added and the reaction stirred at 25 ˚C for 48 hours. The crude product was purified using flash column chromatography (70 mg, 0.2 mmol, 36%).

R_f = 0.14 (1:19 MeOH:EtOAc). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.62 (1H,app d, *J= 1.8 Hz*, CH), 8.52 (1H,app d, *J= 1.8 Hz*, CH), 7.74 (1H,s, CH), 7.31-7.22 (3H, m, 3×CH), 7.065 (2H, d, *J= 7.0 Hz*, 2×CH), 6.16 (1H, br s, NH), 5.21 (2H, s, CH_2), 4.615 (2H, d, *J= 6.0 Hz*, CH_2), 3.74-3.43 (8H, br m, 4×CH_2), 2.41 (3H, s, CH_3), 2.40 (3H, s, CH_3) . **^{13}C NMR** (100 MHz, $CDCl_3$): δ 167.63, 164.95, 150.21, 146.75, 145.85, 142.46, 136.04, 134.76, 134.54, 131.03, 128.86, 127.88, 126.73, 113.29, 66.77, 52.83, 42.53, 40.64, 14.38, 11.12. **HRMS** calculated for $C_{24}H_{27}N_5O_3$ + Na^+ [M + Na^+]: 456.2012 Da. Found [M+ Na^+]: 456.2013 Da (+0.3 ppm).

6.2.24 Ethyl 3,5-dimethyl-1H-pyrazole-4-carboxylate[290]

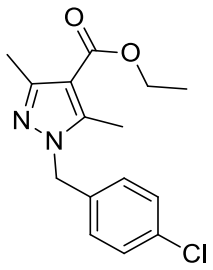


3.28

Ethyl 2-acetyl-3-hydroxybut-2-enoate (2.2 g, 12.5 mmol) was dissolved in anhydrous EtOH (50 mL) under a nitrogen atmosphere and to the reaction mixture hydrazine monohydrochloride (1.1 g, 15 mmol) was added. The reaction was stirred at 25 ˚C for 16 hours and then refluxed for a further hour, which drove the reaction to completion. The solvent was removed *in vacuo* and the crude product triturated with Et_2O (75 mL) resulting in the pure product which appeared as white crystals (1.9 g, 11.5 mmol, 93%).

R_f = 0.4 (1:9 MeOH:DCM) **^{1}H NMR** (400MHz, $CDCl_3$): δ 4.32-4.41 (2H, q, *J*= *7.1 Hz*, CH_2), 2.70 (6H, S, CH_3),1.39 (3H, t, *J*= *7.1 Hz*, CH_3). **^{13}C NMR** (400MHz; $CDCl_3$): δ 162.02, 148.07, 110.86, 61.03, 14.25, 11.90. **HRMS** calculated for $C_8H_{12}N_2O_2$ + H^+ [M+ H^+]: 169.0972 Da. Found [M+ H^+]: 169.0978 Da (-3.99).

6.2.25 Ethyl *N*-(4-chlorobenzyl)-3,5-dimethyl-1H-pyrazole-4-carboxylate[252]

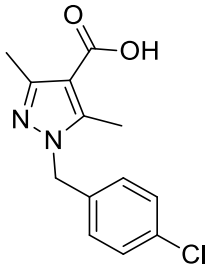


3.29

3.28 (500 mg, 2.9 mmol), 4-chlorobenzylbromide (672.3 mg, 3.3 mmol), K_2CO_3 (452 mg, 3.3 mmol) and cyclohexanone (6 mL) were heated to reflux for 5 hours. The reaction was filtered to remove the K_2CO_3 and the cyclohexanone removed *in vacuo,* the crude reaction residue was taken up into $CHCl_3$ (20 mL) and washed with 1 M NaOH (3 × 15 mL) and distilled water (3 × 15 mL). The organic layer was dried over Na_2SO_4, filtered and the solvent removed *in vacuo*, finally the crude product was purified using flash column chromatography (526.5 mg, 1.8 mmol, 60%).

R_f = 0.74 (6:4 EtOAc:Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ. 7.28 (2H, d, *J=8.4 Hz*, 2×CH), 7.03 (2H, d, J=8.4Hz, 2×CH), 5.20 (2H, s, CH_2), 4.31 - 4.26 (2H, q, *J=7.1Hz*, CH_2), 2.44 (3H, s, CH_3), 2.43 (3H, s, CH_3), 1.35 (3H, t, *J=7.1Hz*, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ. 164.48, 150.80, 144.10, 134.68, 133.75, 129.03, 128.08, 110.29, 59.67, 52.11, 14.39, 14.35, 11.28. **HRMS** calculated for $C_{15}H_{18}ClN_2O_2 + H^+$ [M+ H^+]: 293.1051 Da. Found [M+ H^+] 293.1054 Da (-0.88 ppm). **Elemental analysis** calculated (%) for $C_{15}H_{18}ClN_2O_2$: C 61.54, H 5.85, N 9.57. Found C 61.49, H 5.84, N 9.12.

6.2.26 *N*-(4-chlorobenzyl)-3,5-dimethyl-1H-pyrazole-4-carboxylic acid[334]

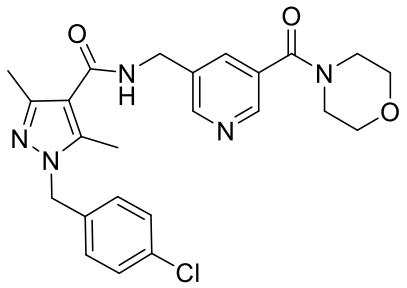


3.30

3.29 (150 mg, 0.5 mmol) was dissolved in MeOH (5 mL) to this solution 1 M NaOH (2.1 mL, 2.1 mmol) was added and the reaction refluxed for 3 hours. The solvent was removed *in vacuo* and the resultant off white solid was dissolved in distilled water (10 mL), the aqueous layer was then washed with Et_2O (3 × 10 mL). The aqueous layer was acidified with the drop wise addition of concentrated H_2SO_4 at 3 ˚C precipitating the carboxylic acid product. The product was extracted into EtOAc (3 × 10 mL), the organic fractions combined, dried over Na_2SO_4. The solvent was removed *in vacuo* yielding the product as a tan solid (124.8 mg, 0.5 mmol, 92%).

R_f = 0.63 (1:9 MeOH:DCM) **^{1}H NMR** (400 MHz, MeOD): δ 7.23 (2H, d, J=8.5 Hz, 2×CH), 6.99 (2H, d, J=8.5 Hz, 2×CH), 5.17 (2H, s, CH_2), 2.36 (3H, s, CH_3), 2.29 (3H, s, CH_3). **^{13}C NMR** (100 MHz, MeOD): δ167.38, 152.36, 146.36, 136.56, 134.71, 129.96, 19.47, 111.28, 52.57, 14.57, 11.30. **HRMS** calculated for $C_{13}H_{14}N_2O_2 + H^+$ [M + H^+]: 265.0744 Da. Found [M+ H^+]: 265.0742 Da (-0.7 ppm).

6.2.27 Inhibitor 5[288,332]

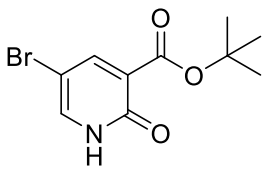


Inhibitor 5

3.30 (113.5 mg, 0.4 mmol) and HATU (3286.0 mg, 0.9 mmol) were stirred in anhydrous DMF (1.1 mL) at 25 ˚C for 20 minutes, after this time K_2CO_3 (237.7 mg, 1.7 mmol) was added and the reaction stirred for a further 30 minutes at 25 ˚C. **3.21** (95.3 mg, 0.4 mmol), dissolved in anhydrous DMF (1 mL) was then added and the reaction stirred at 25 ˚C for 60 hours. The solvent was removed *in vacuo*, and the dark orange oil residue was taken up into EtOAc (10 mL) and washed with saturated $NaHCO_3$ (3 × 20 mL) and distilled water (3 × 20 mL). The organic layer was dried over anhydrous Na_2SO_4, filtered and the solvent removed *in vacuo,* the crude product was purified using flash column chromatography (68 mg, 0.1 mmol, 33.9%).

R_f = 0.10 (1:19 MeOH:EtOAc). **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.65 (1H,s, CH), 8.55 (1H, s, CH), 7.77 (1H, s, CH), 7.30 - 7.28 (2H, d, *J= 8.3 Hz*, 2×CH), 7.04 (2H, d, *J= 8.3 Hz, 2*×CH), 6.12 (1H, br s, NH), 5.20 (2H, s, CH_2), 4.65 (2H, d, *J= 6.0 Hz*, CH_2), 3.77 - 3.45 (8H, br m, 4×CH_2), 2.43 (3H, s, CH_3), 2.42 (3H, s, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 167.62, 164.81, 150.21, 146.78, 146.00, 142.48, 134.70, 134.57, 134.53, 133.84, 131.06, 129.06, 128.16, 113.41, 66.78, 52.11, 48.22, 40.65, 14.40, 11.09. **HRMS** calculated for $C_{24}H_{26}N_5O_3 + Na^+$ [M + Na^+]: 490.1622 Da. Found [M + Na^+]: 490.1611 Da (-2.2 ppm).

6.2.28 *tert*-Butyl-5-bromo-2-oxo-1,2-dihydropyridine-3-carboxylate[328]

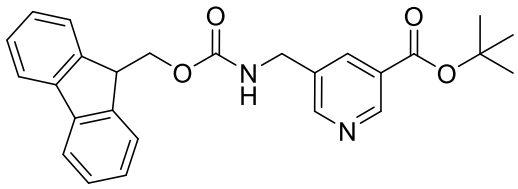


i

To a stirring solution of 5-bromo-2-oxo-1,2-dihydropyridine-3-carboxylic acid (500 mg, 2.3 mmol) in $CHCl_3$ (5 mL), 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDCI) (393.5 mg, 2.5 mmol), DMAP (309.5 mg, 2.5 mmol) and t-butanol (0.8 mL, 9.2 mmol) was added. The reaction was left stirring at 25 ˚C for 72 hours and purified using flash column chromatography (204.5 mg, 0.7 mmol, 32.5%).

R_f = 0.52 (50:50 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.34(1H, s, CH), 8.18(1H, s, CH), 1.61(9H, s, 3×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 166.93, 164.00, 153.32, 142.36, 111.80, 108.82, 84.75, 28.09. **HRMS** calculated for $C_{10}H_{12}BrNO_3$ + H^+ [M + H^+] : 274.0073 Da. Found [M + H^+] 274.0073 Da (-0.1 ppm).

6.2.29 *tert*-Butyl-5-(9H-fluorenylmethoxycarbonyl)-amino methyl nicotinate[335]

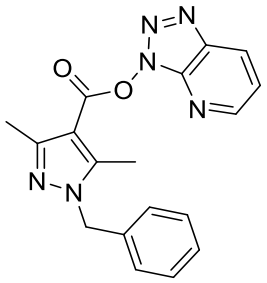


ii

To a solution of **3.17** (100 mg, 0.5 mmol) in anhydrous MeOH (3.6 mL) at 0 °C Nickel (ii) chloride hexahydrate (11.5 mg 0.05 mmol) followed by sodium borohydride ($NaBH_4$) (129.5 mg, 3.4 mmol) were added in portions over 30 minutes. The reaction was warmed to 25 ˚C and left stirring for 4 hours with monitoring using TLC after which time there appeared to be no starting material present. Fluorenylmethyloxycarbonyl chloride (Fmoc-cl) (253 mg, 1.0 mmol) was then added and the reaction left stirring for 16 hours, upon complete reaction of the amine intermediate the crude reaction was purified using flash column chromatography to afford the product which appeared as an orange oil. (8.6 mg, 0.04%).

R_f = 0.38 (1:1 EtOAc: Hex) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 9.04(1H, s, CH), 8.64(1H, s, CH), 8.16 (1H, s, CH), 7.74(2H, d, *J = 6.9 Hz*, 2×CH), 7.56(2H, d, *J = 6.7 Hz*, 2×CH), 7.38 (2H, t, *J = 6.8 Hz*, 2×CH) 7.31- 7.25 (2H, m, 2×CH), 4.42 (4H, app dd, *J= 14.5, 5.6 Hz*, 2×CH_2), 4.19 (1H, t, *J= 5.9 Hz*, CH), 2.20(1H, s, NH), 1.59 (9H, s, 3×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 164.57, 156.92, 152.47, 150.09, 144.15, 141.68, 136.50, 134.43, 128.10, 127.45, 125.34, 120.37, 82.69, 67.28, 58.63, 47.56, 28.50. **HRMS** calculated for $C_{26}H_{26}N_2O_4$ + H^+ [M + H^+]: 453.1790 Da. Found [M + H^+] 453.1789 Da (-0.3 ppm).

6.2.30 3H-[1,2,3]triazolo[4,5-b]pyridin-3-yl *N*-benzyl-3,5-dimethyl-1H-pyrazole-4-carboxylate[288,332]

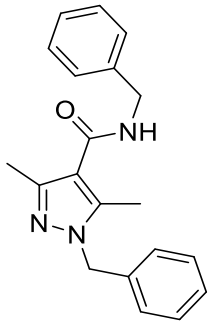


iii

3.25 (67.4 mg, 0.3 mmol), was dissolved in anhydrous DMF (0.7 mL), to this solution was added HATU (201.1 mg, 0.5 mmol) and the solution left to stir at 25 ˚C for 30 minutes, after this time DIPEA (203.8 µL, 1.2 mmol) was added and the reaction left for a further 30 minutes at 25˚C. **ii** (64.7 mg, 0.3 mmol) dissolved in anhydrous DMF (0.5 mL) was then added and the reaction left to stir at 25 ˚C for 24 hours. After this time the solvent was removed *in vacuo*, the residue partitioned between EtOAc (5 mL) and $NaHCO_3$ (5 mL), the organic layer was washed a further 3 × $NaHCO_3$ (10 mL) and dried over Na_2SO_4. The solvent removed *in vacuo* to afford the crude product which appeared as a dark orange solid (131 mg).

R_f =0.80 (1:19 MeOH:DCM) **^{1}H NMR** (400 MHz, $CDCl_3$): δ 8.73 (1H, dd, *J= 4.4, 1.0 Hz*, CH), 8.44 (1H, dd, *J= 8.4, 1.1 Hz*, CH), 7.46 - 7.42 (1H, m, CH), 7.38 - 7.29 (3H, m, 3×CH), 7.15 (2H, app d, *J = 7.1 Hz*, 2×CH), 5.33 (2H, s, CH_2), 2.60 (3H, s, CH_3), 2.56 (3H, s, CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 159.75, 152.09, 151.74, 146.99, 141.00, 135.26, 135.08, 129.54, 129.04, 128.21, 126.80, 120.79, 104.90, 53.36, 14.67, 11.91. **HRMS** calculated for $C_{18}H_{16}N_6O_2$ + H^+ [M + H^+]: 371.1232 Da. Found [M + H^+] 371.1222 Da (-2.8 ppm).

6.2.31 *N*,1-dibenzyl-3,5-dimethyl-1H-pyrazole-4-carboxamide[288,332]



iv

iii (87.6 mg, 0.3 mmol) was dissolved in anhydrous THF (1.2 mL) to this solution benzyl amine (41.2 µL, 0.7 mmol), DIPEA (131.1 µL, 0.7 mmol) and anhydrous DMF (5 drops) were added. The reaction was stirred at 25 ˚C for 96 hours, after this time the solvent was removed *in vacuo*. The residue partitioned between EtOAc (5 mL) and $NaHCO_3$ (5 mL), the organic layer was washed a further 3× $NaHCO_3$ (10 mL) and dried over Na_2SO_4. The solvent removed *in vacuo* to afford the crude product which appeared as an off white solid (52.8 mg, 0.2 mmol).

^{1}H NMR (400 MHz, $CDCl_3$): 7.34-7.25 (8H, m, 8× ar CH), 7.08 (2H, d, *J=7.1 Hz*, 2×ar CH), 5.21 (2H, s, CH_2), 4.59 (2H, d, *J= 5.6 Hz*, CH_2) 2.42 (6H, app d, *J= 3.5 Hz*, 2×CH_3). **^{13}C NMR** (100 MHz, $CDCl_3$): δ 164.68, 145.85, 142.13, 138.56, 136.22, 128.82, 128.78, 127.81, 127.69, 127.50, 126.73, 113.86, 52.76, 43.53, 14.25, 11.03. **HRMS** calculated for $C_{20}H_{21}N_3O$ + H^+ [M + H^+]: 342.1582 Da. Found [M + H^+]: 342.1576 Da (-1.8ppm).

6.3 Biological Materials and Methods

6.3.1 Materials

6.3.1.1 Media

General Solutions

There were two types of water used in the biophysical analysis of the compounds within this thesis, within this specific chapter they are listed as being either RO water, or MilliQ water.

RO water is that which has been de-ionised and distilled, it was used for those purposes such as media solutions whereby sterilisation by autoclave is a general method involved in ensuring aseptic materials.

MilliQ water was obtained by the use of a “Synergy Water Purification system (Millipore) fitted with a SynergyPak® 1 cartridge and a MilliPak-20 Express system”, the purity of this water meant that it was used for buffer solutions. Organic solvents such as Ethanol and Methanol were obtained from Fisher Scientific and any reagents used in the methods were supplied by Sigma-Aldrich at general laboratory grade. All pH measurements were made at 25˚C and adjusted using 1M HCl of 1M NaOH.

SOC Media

Make up to 200mL using RO water do not add glucose until autoclaved as heat degrades it, sterile filter (22µM Millex® HA syringe-driven filter unit) into 50mL falcon tubes for storage.

- RO water (200 mL)
- Tryptone (20 g)
- Yeast extract (5 g)
- NaCl (10 mM)
- KCl (25 mM)
- MgCl2.6H2O (10 mM)
- pH 7.0
- Autoclave
- Glucose (20 mM)

LB Broth

Autoclave before the addition of sterile filtered (Millex® HA syringe-driven filter unit) kanamycin (adding prior to this will render the antibiotic ineffective).

- LB broth granules (25 g/1 L RO water)
- Kanamycin (32 µg/mL)

LB Agar

Autoclave before the addition of sterile filtered (Millex® HA syringe-driven filter unit) kanamycin (adding prior to this will render the antibiotic ineffective).

- LB Agar granules (37 g/1L RO water)
- Kanamycin (32 µg/mL)

Minimal Media A

- RO water
- Na_2HPO_4 (88 mM)
- KH_2PO_4 (55 mM)
- pH= 7.2

Minimal Media B

- RO water
- Thiamine.HCl (30 µM)
- $CaCl_2.2H_2O$ (135 µM)
- $MgSO_4.7H_2O$ (1 mM)
- NH_4Cl (20 mM)
- Glucose (20 mM)

6.3.1.2 Buffer Solutions

HisTrap A (loading/running buffer)

- MilliQ water (1 L)
- Tris/HCl (50 mM)
- NaCl (500 mM)
- pH 7.5
- Filter (Millipore) and de-gas

HisTrap B (elution buffer)

- MilliQ water (1 L)
- Tris/HCl (50 mM)
- NaCl (500 mM)

- Imidazole (500mM, Fluka)
- pH 7.5
- Filter (Millipore) and de-gas

Gel Filtration Buffer

- MilliQ water (1 L)
- Tris/ HCl (50 mM)
- NaCl (150 mM)
- pH 7.5

15% PAGE Resolving Gel (6.3.2.6 experimental procedure)

- MilliQ (2.4 mL)
- 30% bis-acrylamide (5 mL)
- Tris HCl pH 8.8 (2.5 mL)
- 10% (w/v) sds (100 µL)
- TEMED (7.5 µL)
- 10% APS (75 µL)

4% PAGE Stacking Gel (6.3.2.6 experimental procedure)

- MilliQ (6.1 mL)
- 30% bis-acrylamide (1.3 mL)
- Tris HCl pH 6.8 (2.5 mL)
- 10% (w/v) sds (100 µL)
- TEMED (7.5 µL)
- 10% APS (75 µL)

10×SDS-PAGE Running Buffer (6.3.2.6 experimental procedure)

- MilliQ water (1 L)
- Tris HCL (250 mM)
- Glycine (1.92 M)
- SDS (1%)

PAGE Coomassie G250 Stain (6.3.2.6 experimental procedure)

- Coomassie Blue G250 (0.1%)
- MeOH (45%)
- MilliQ (44.9%)
- Acetic Acid (10%)

PAGE De-Stain (6.3.2.6 experimental procedure)

- MeOH (45%)
- MilliQ (45%)
- Acetic Acid (10%)

NMR Buffer (proteins)

- Tris HCl (50 mM)
- NaCl (50 mM)
- $CaCl_2$ (5 mM)
- pH 6.4
- Filter (Millipore) and de-gas

ITC Buffer

- Tris HCl (50 mM)
- NaCl (50 mM)
- $CaCl_2$ (5 mM)
- pH 7.5
- Filter (Millipore) and de-gas

Fluorescence Spectrophotometry Buffer

- Tris HCl (50 mM)
- NaCl (50 mM)
- $CaCl_2$ (5 mM)
- pH 7.5
- Filter (Millipore) and de-gas

6.3.2 Methods

6.3.2.1 General Equipment

The methods detailed below describe the use of specific volumes, to ensure accuracy all measurements under 10 mL were carried out using Gilson single channel pipettes (Pipetman) anything above 10 mL was carried out using graduated measuring cylinders. Specific machinery has been detailed, including the make and model and unless otherwise stated is available for use within the Institute of Integrative Biology, Department of Biochemistry, University of Liverpool.

6.3.2.2 Transformation

NCS1 was obtained from an expression construct of *Rattus Norvegicus,*[18] it was sub-cloned from a p-GEX-6p plasmid expressing NCS1 [336] into a pETM-11 vector

and subsequently transformed into BL21 DE3 *E-coli* competent cells (Novagen). The NCS1 construct in pETM-11 vector was obtained from another member within the lab, however the transformation method is detailed here: 2 µL of the NCS1 containing plasmid was added to 50 µL of BL21 DE3 E-coli competent cells (Novagen), the solution was mixed and incubated at 0 °C for 30 minutes. Following this the cells were incubated at 42 °C for 30 seconds and then cooled back to 0 °C for a further 5 minutes. After this time 200 µL of a pre-prepared SOC media was added and the solution incubated with shaking at 180 rpm for an hour at 37 °C. 200 µL of this solution was subsequently pipetted under aseptic conditions onto a kanamycin containing agar plate, spread evenly over the surface and the plate allowed to dry at 25 °C for around 20 minutes. The agar plate was then incubated at 37 °C for 16 hours without shaking, it was allowed to cool to 25 °C for around 30 minutes and finally was sealed with Parafilm® and stored at 4 °C for a maximum of two weeks.

6.3.2.3 NCS1 expression in Minimal Media

A starter culture of Luria Broth (LB) with kanamycin of a final concentration 32 µg/ mL, was inoculated with a single colony from the LB agar/ kanamycin transformed plate of NCS1 and incubated at 37 °C with shaking at 180 rpm for 7 hours. A preparation of Minimal Media A (MMA) and B (MMB) solutions were produced. MMA solutions were autoclaved and MMB syringe-filtered through a 0.22 µM Millex® Millipore HA filter unit. The minimal media solutions (500 mL MMA with 6.5 mL MMB) were supplemented with 32 µg/mL kanamycin, to 10 mL of this solution 100 µL of the starter culture was added and incubated overnight at 37 °C with shaking. This cell suspension was then added to the minimal media solution and an initial optical density (OD_{600}) reading was taken (ca 0.05 - 0.1). The cells were grown at 37 °C for 6 hours to reach an OD_{600} = 0.8; this was followed by induction of expression through the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration 1 mM) and then incubated at 18 °C for 16 hours with shaking. Centrifugation of the cells was carried out at 1519 g for 20 minutes at 4 °C and the subsequent supernatant was discarded. The cells were harvested and re-suspended in HisTrap buffer A, to this solution a complete EDTA free protease inhibitor tablet (Roche) pre-dissolved in 1 mL of HisTrap A buffer was then added, the mixture was flash frozen using liquid nitrogen and stored at -80 °C until purification.

6.3.2.4 NCS1 expression in LB Media

A 50mL solution of Kanamycin selective LB solution was inoculated with a single colony from the transformed plate of NCS1, it was subsequently incubated at 37 °C with shaking at 180 rpm for 16 hours (overnight). The cells were centrifuged at 3,000

g and a temperature of 4 ºC for 14 minutes, the pellet was re-suspended in a total volume of 5 mL kanamycin selective LB. The cell suspension was then added to a 1 L solution of kanamycin selective LB in a 2 L conical flask and an initial optical density (OD_{600}) reading was taken (normally around 0.05 - 0.1). The cells were incubated at 37 ºC until they reached an OD_{600} = 0.8 which normally occurred within 4 hours. Following this the cells were induced with the addition IPTG at a final concentration 1 mM and then incubated at 18 ºC for 16 hours with shaking. The induced cells were centrifuged at 1519 g for 20 minutes at 4 ºC and the subsequent supernatant was discarded. The cells were harvested and re-suspended in HisTrap buffer A, to this solution a complete EDTA free protease inhibitor tablet (Roche) pre-dissolved in 1 mL of HisTrap A buffer was then added, the mixture was stored at -80 ºC until purification.

6.3.2.5 Protein Purification Methods

6.3.2.5.1 Cell Lysis

The frozen cells stored at -80 ˚C were thawed gently on ice and treated with ribonuclease1 from bovine pancreas (sigma) to promote nucleic acid breakdown, a final concentration 12 µg/mL and the solution incubated for 10 minutes at 4°C. The cells were lysed mechanically at 0 ˚C in 10 mL portions using a Stansted 'Pressure Cell' Homogeniser (SFP Ltd), at 1000 P.S.I. This was followed by centrifugation of the cells at 47,813 g for 30 minutes at 4 ºC to pellet any insoluble cell debris. The resultant supernatant was decanted and sterile filtered through a 0.45 µM Millex® HA syringe-driven filter unit in preparation for the subsequent purification using liquid chromatographic techniques.

6.3.2.5.2 Ni^{2+} Affinity Chromatography

The His-tagged protein solution was then purified using Ni^{2+} affinity chromatography using a HisTrapFF 5 mL column (GE Healthcare) on an ÄKTA purifier system (GE Healthcare). The column was been washed with MilliQ and then pre-equilibrated with 4 column volumes of HisTrap buffer A and the imidazole containing buffer B at a flow rate of 2.5 mL/min. The filtered cell supernatant was loaded onto the column at a flow rate of 2.5 mL/min, the His-tagged protein that bound to the column was then eluted by increasing the gradient of imidazole at a flow rate of 2.5 mL/min. Eluent was fractionated with a sample size of 5 mL, analysis of the fractions was conducted using Sodium Disulfate Polyacrylamide (SDS) polyacrylamide gel electrophoresis (SDS PAGE 6.3.2.3). Those observed as being protein containing fractions (observed as containing the desired gel band) were pooled in preparation for concentration.

6.3.2.5.3 Concentration and Buffer exchange of protein samples

Following Ni^{2+} affinity chromatography the combined protein fractions were concentrated using an Amicon Ultra Centrifugal Filter at 3,500 g which had been previously washed with MilliQ and equilibrated in HisTrap buffer A. The protein solution was concentrated to the desired final volume, the eluent that passed through the membrane was analyzed using SDS PAGE to ensure that none of the protein of interest had passed through. The concentration units were subsequently washed with MilliQ and 20% EtOH by centrifugation at 3,500 g after which they were then stored for reuse. The protein solution was then buffer exchanged back into a non-imidazole containing buffer HisTrap A using a desalting column (Sephadex G-25 Medium Gravity-Flow PD-10 (GE Healthcare)) that had been washed with MilliQ and pre equilibrated with HisTrap A. The process of column equilibration involved washing the column with 25 mL of MilliQ followed by 25 mL of HisTrap A, the protein was subsequently loaded onto the column in a volume of 2.5 mL and any flow through at this time was discarded. The sample was then eluted by the addition of 3.5 mL HisTrap A, this process was repeated until all the protein sample had been buffer exchanged as the maximum volume of protein that can be added to the column at any one time is 2.5 mL. The protein sample was collected for the subsequent cleavage process and the PD-10 column was washed with 25 mL of MilliQ and 25mL 20% EtOH, it was stored in 20% EtOH at room temperature.

6.3.2.5.4 Tobacco Etch Virus (TEV) protease cleavage of Histidine tag and reverse purification

A 20:1 protein: protease ratio of Tobacco Etch Virus (TEV) was added to the protein solution after the buffer exchange process, the solution was left inverting at 4 °C overnight. The cleavage solution was then passed back through the Ni^{2+} affinity column at 2 mL/min, any un-cleaved protein would remain on the column and the cleaved protein is eluted in HisTrap buffer A. The cleaved tag or un-cleaved protein that had remained on the column was then eluted in HisTrap buffer B.

6.3.2.5.5 Size exclusion Chromatography

This cleaved NCS1 after the reverse purification protocol was then concentrated to 5 mL and size exclusion gel filtration chromatography was conducted using a Hi Load 26/60 Superdex 75 column (Amersham Biosciences). The column was connected to an AKTApurifier10 FPLC system and was initially washed with MilliQ before equilibration with gel filtration buffer containing 50 mM tris/ HCl, 150 mM NaCl at pH 7.5. The protein was manually injected using a 10 mL injection loop which was connected to the FPLC, loading of the sample onto the column was then carried out

at a flow rate of 3 mL/min. Increasing volumes of gel filtration buffer eluted the protein into fractions that were analyzed using SDS PAGE, those fractions containing NCS1 at the desired molecular weight were pooled and then flash frozen using liquid nitrogen and stored at -80 °C.

6.3.2.6 Sodium Disulfate Polyacrylamide Gel Electrophoresis (SDS PAGE)

SDS PAGE was used to monitor the expression and purification process of NCS1 and to assess the purity of the final protein sample using a method developed by Laemmli.[337] Full casting required a 15% resolving gel (R) prepared using 2.4 mL of water, 5 mL 30% bis-acrylamide, 2.5 mL tris HCl buffer pH 8.8 and 100 µL of 10 × SDS solution, to this solution 7.5 µL of tetramethylethylenediamine (TEMED) was added along with 75 µL of a 10% solution of ammonium persulfate (APS). The solution was pipetted into a cast of two glass plates and a small volume no greater than 500 µL of butanol was pipetted on top of the R gel to remove any air trapped. The gels were left to polymerize for 20 minutes, during which time 10 mL of a 4% stacking gel (S) was prepared with 6.1 mL of water, 1.3 mL 30% bis-acrylamide, 2.5 mL tris/ HCl buffer pH 6.8 , 100 µL 10× SDS. Once the R layer had polymerized the butanol layer was removed, 7.5 µL of TEMED was added with 75 µL of a 10% solution of APS to the S layer solution this was pipetted onto the R gel layer and either a 10 or 15 sample well comb was inserted. The S gel was then allowed to polymerize for a further 20 minutes.

The gel samples were prepared by diluting the protein sample with up to a volume of 20 µL of water and 20 µL of 2× running buffer was then added. The samples were then heated to 100 °C and 7.5 µL of each were loaded into individual wells on the gel, 7.5 µL of a Sigma low molecular weight marker (sigma m3913) was also added to a separate lane.

The gels were removed from the glass cast and washed with RO water, before staining with a stain solution containing Coomassie blue G250 (0.1%), MeOH (45%), MilliQ water (44.9%) and acetic acid (10%) was added to the gel which was then incubated and left to stain with agitation at room temperature for 10 minutes. The stain was decanted off and the gels were again washed with water, de-stain solution (MeOH (45%), MilliQ water (45%) and acetic acid (10%)) was then added and the gels left for 12 hours at 20-25°C with agitation to remove any remaining stain. The gels were then visualised using an image scanner (Image III GE Healthcare).

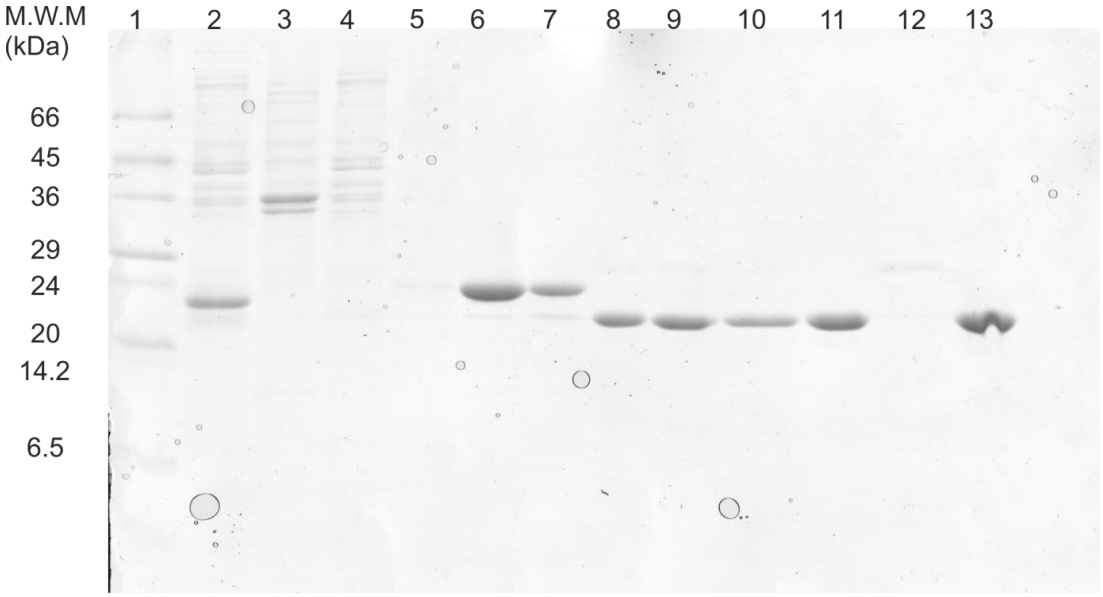


Figure 6.1- SDS-Page Electrophoresis Gel of NCS1- MM unlabeled protein preparation of NCS1 monitored using gel electrophoresis; **Lane1:** Low molecular weight marker, **Lane 2**: Supernatant, **Lane 3**: Pellet, **Lane 4**: HisTrap fraction 2, **Lane 5**: HisTrap fraction 6,**Lane 6**: HisTrap fraction 18, **Lane 7**: Pre TEV cleavage, **Lane 8**: Post TEV cleavage, Pre buffer exchange, **Lane 9**: Post buffer exchange, **Lane 10**:Conc down sample, **Lane 11**: 1st His elute, **Lane 12**: 2nd His elute, **Lane 13**: Gel filtration fraction 18.

6.3.2.7 NMR

General

All NMR spectra were collected at 25 ˚C unless otherwise stated, on either a Bruker AVANCE II+ 600 MHz Ultrashield or 800 MHz US^2 spectrometer, equipped with (TCI) triple-resonance cryoprobes. The NMR samples were prepared to a volume of 550µL including 10% ^2H_2O and the NMR buffer used was composed of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$, pH 6.4. Data collection used Topspin 3.1 (Bruker, UK) and sample quality was assessed using comparison to earlier spectra collected ^{15}N and ^{13}C NCS1.

6.3.2.7.1 Backbone Assignment

A 0.967 mM sample of ^1H ^{15}N ^{13}C NCS1 was exchanged into NMR buffer using a desalting column (Sephadex G-25 Medium Gravity-Flow PD-10 (GE Healthcare)). The method for protein back bone assignment has been described previously (2.3.1) and the methodology utilises the following experiments: 2-D ^1H ^{15}N Heteronuclear Single Quantum Coherence spectra (HSQC) and 3-D triple resonance experiments HCCH TOCSY, C_bC_a(CO)NH, HNC_aC_b, HNCO, HN(C_a)CO, the types of experiments used has been extensively reviewed.[255–261,338] Data collection was conducted using standard practises and data was analysed using the Collaborative Computational Project for NMR (CCPN) Analysis software programme.[339]

6.3.2.7.2 First Generation Ligand Screening.

A 29.95 mM stock solution for each inhibitor was prepared by dissolving the appropriate mass of inhibitor in the required volume of dimethyl sulfoxide (DMSO). A 0.5 mM stock solution of ^{1}H ^{15}N NCS1 was prepared in 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$, pH 6.4 (NMR buffer), this was diluted to 97.9 µM and the required ligand stock added to yield a final NCS1: ligand ratio of 1:1 and 1:10 for Inhibitor 1, 1:10 for Inhibitor 2 and Inhibitor 3 along with 10% $^{2}H_2O$ (50 µL). To ensure that any perturbations seen upon addition of the ligand were due to binding and not an artefact of DMSO binding to the protein, the appropriate volume of DMSO for each inhibitor were added to the apo sample of ^{1}H ^{15}N NCS1 and the spectrum collected for comparison.

6.3.2.7.3 Second Generation Ligand Binding Screening.

A 100 mM stock solution for Inhibitor was prepared by dissolving the appropriate mass of inhibitor in the required volume of DMSO as with the first generation of inhibitors. A 96.5 µM stock solution of ^{1}H ^{15}N NCS1 was prepared in NMR buffer, the required volume of protein to achieve a concentration of 50 µM in a total final volume of 550 µL was added to an eppendorf along with 10% $^{2}H_2O$ (50 µL) and the required volume of inhibitor stock solution to yield a final concentration of 500 µM or 1 mM respectively. Therefore the ligand to protein ratio was either 10:1 or 20:1. As with the first generation of inhibitors to ensure that any perturbations seen upon addition of the ligand were due to binding and not an artefact of DMSO binding to the protein, the appropriate volume of DMSO for each inhibitor were added to the apo sample of ^{1}H ^{15}N NCS1 and the ^{1}H NMR spectrum collected for comparison.

6.3.2.7.4 Solubility comparison experiment

From the stock solutions of the Inhibitors made previously (6.3.2.7.2 and 6.3.2.7.3) the required volume of inhibitor was added to an eppendorf to achieve a final concentration of 500 µM in a final volume of 550 µL, 100 µM TSP was added from a stock solution of 1 mM along with 10% $^{2}H_2O$ (50 µL). A ^{1}H 1-D NMR spectrum was collected, the reference TSP height was adjusted so that it was consistent within each experiment and the inhibitor peaks interpreted accordingly.

6.3.2.8 Small Molecule Compound Library Protocols

6.3.2.8.1 Initial Screen

Each of the 28 compounds were dissolved in the required volume of DMSO to achieve a stock concentration of 100 mM. A 96.5 μM stock solution of ^{1}H ^{15}N NCS1 was prepared in NMR buffer. For each compound sample the required volume of protein to achieve a concentration of 20 μM was subsequently added in an eppendorf containing the required volume of small molecule to achieve a 1 mM concentration along with 10% $^{2}H_2O$ (50 μL). ^{1}H 1-D NMR spectra and saturation transfer difference (STD) experiments were collected.[307,308] This data was compared to a number of standards which include the same experiments collected on the following: a sample of 1 mM compound along with 10% $^{2}H_2O$ (50 μL) in NMR buffer and a 20 μM sample of NCS1 along with 10% $^{2}H_2O$ (50 μL) and the required volume of DMSO to achieve the same concentration as that which is added with the small molecules.

6.3.2.8.2 Secondary Testing

To the samples of those small molecules which appeared to display an interaction with the protein a further aliquot of NCS1 was added to achieve the final concentration of 50 μM. ^{1}H, STD, waterLOGSY and ^{1}H ^{15}N HSQC spectra were collected and compared to the standards which included the same experiments collected on the following samples: 1mM compound along with 10% $^{2}H_2O$ (50 μL) in NMR buffer and a 50 μM sample of NCS1, 10% $^{2}H_2O$ (50 μL) and the required volume of DMSO to achieve the same concentration as that which is added with the small molecules.[292,293]

6.3.2.8.3 D2 Competition Experiments

D2R peptide was dissolved in the required volume of MilliQ water to achieve a stock concentration of 1 mM at pH of 6.4, it was aliquoted into 28.89 μL volumes in a series of eppendorf's and freeze dried for 16 hours. The dried D2R peptide was then re-suspended in the NMR sample from 6.3.2.8.2 to achieve a 50 μM final concentration of the peptide. ^{1}H 1-D STD, ^{1}H 1-D waterLOGSY and ^{1}H ^{15}N HSQC spectra were collected and compared to the standards which included the same experiments collected on the following samples: 1 mM compound along with 10% $^{2}H_2O$ (50 μL) in NMR buffer, a 50 μM sample of NCS1, 10% $^{2}H_2O$ (50 μL), 50 μM D2R peptide and the required volume of DMSO to achieve the same concentration as that which is added with the small molecules and a 50 μM sample of D2R peptide in NMR buffer.

6.3.2.9 Isothermal Titration Calorimetry (ITC) Protocol

ITC experiments were conducted at 25 ^{o}C and were run on an iTC_{200} Microcalorimeter (MicroCal), this has a 200 µL cell capacity and syringe volume 40 µL. Due to solubility issues with both the protein and inhibitors all experiments of the first generation were run in reverse, meaning the ligand was held within the cell and the protein titrated in from the syringe, the concentration ranges for protein and inhibitor are as follows unlabelled NCS1: 1 mM - 400 µM, Inhibitor 1; 600 µM - 100 µM. The preparation of NCS1 and inhibitors for all experiments used a salt buffer of 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ at pH 7.5. Data collected was analysed using Origin®7 software and a full description of the calorimeter and the analysis software can be found in 2.3.5. Second generation experiments were conducted at 25 ^{o}C and 10 ^{o}C in the classical method of titrating the ligand into the cell containing NCS1, the concentration of protein and inhibitor are as follows unlabelled NCS1: 400 µM, Inhibitor 5; 8 mM.

6.3.2.10 Fluorescence Spectrophotometry

All fluorescence data of the first generation of inhibitors was collected by Dr Liam Dorr, second generation inhibitor data collected by myself at The University of Liverpool on a Cary Eclipse Fluorescence spectrophotometer (Agilent technologies). The raw data of both generations of inhibitors was then analysed as part of this project/ thesis in Origin7 (OriginLab, Northampton, MA). A stock solution of unlabelled NCS1 45 µM was prepared in fluorescence buffer (50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5); and aliquoted into 17 samples at a final concentration of 1 µM v/v (200 µL). For the individual single point reads of the tryptophan fluorescence assay, each inhibitor was added in the required volume from a single stock solution of 100 µM in 100% DMSO, to the already aliquoted NCS1 samples contained in eppendorf's. The concentration range for each inhibitor over the 17 samples was 0-90 µM. This process was repeated for all inhibitors, each individual assay sample was transferred to a quartz 16.160-F/Q/10 cuvette (Starna Scientific) and excited at a wavelength (λ) of 280 nm, with the excitation slit width 5nm and the emission slit width 20 nm.

References

[1] M. Amici, A. Doherty, J. Jo, D. Jane, K. Cho, G. Collingridge, et al., Neuronal calcium sensors and synaptic plasticity., Biochem. Soc. Trans. 37 (2009) 1359–63.

[2] M. Ikura, Calcium binding and conformational responce in EF-hand proteins, 21 (1996) 14–17.

[3] R. D Burgoyne, The neuronal calcium-sensor proteins., Biochim. Biophys. Acta. 1742 (2004) 59–68.

[4] S. Lim, T. Strahl, J. Thorner, J.B. Ames, Structure of a Ca2+-myristoyl switch protein that controls activation of a phosphatidylinositol 4-kinase in fission yeast., J. Biol. Chem. 286 (2011) 12565–77.

[5] K.-H. Braunewell, A.J.K. Szanto, Visinin-like proteins (VSNLs): interaction partners and emerging functions in signal transduction of a subfamily of neuronal Ca2+-sensor proteins, Cell Tissue Res. 335 (2009) 301–316.

[6] W.F. An, M.R. Bowlby, M. Betty, J. Cao, H.-P. Ling, G. Mendoza, et al., Modulation of A-type potassium channels by a family of calcium sensors, Nature. 403 (2000) 553–556.

[7] A. V Tzingounis, M. Kobayashi, K. Takamatsu, R.A. Nicoll, Hippocalcin gates the calcium activation of the slow afterhyperpolarization in hippocampal pyramidal cells, Neuron. 53 (2007) 487–493.

[8] T. Tsujimoto, A. Jeromin, N. Saitoh, J.C. Roder, T. Takahashi, Neuronal calcium sensor 1 and activity-dependent facilitation of P/Q-type calcium currents at presynaptic nerve terminals., Science. 295 (2002) 2276–2279.

[9] C.N. Johnson, S.M. Damo, W.J. Chazin, EF-Hand Calcium-Binding Proteins, in: eLS, John Wiley & Sons, Ltd, 2001.

[10] D.W. O'Callaghan, B. Hasdemir, M. Leighton, R.D. Burgoyne, Residues within the myristoylation motif determine intracellular targeting of the neuronal Ca2+ sensor protein KChIP1 to post-ER transport vesicles and traffic of Kv4 K+ channels., J. Cell Sci. 116 (2003) 4833–45.

[11] T.. A.J.B.. H.T.S.. S.L.. I.M. Tanaka, Sequestration of the membrane-targeting myristoyl group of recoverin in the calcium-free state., Nature. 376 (1995) 444–447.

[12] R.. B.G.. G.M.. P.K.. S.M.C. Stephen, Stabilizing function for myristoyl group revealed by the crystal structure of a neuronal calcium sensor, guanylate cyclase-activating protein 1., Structure. 15 (2007) 1392–1402.

[13] S. Pandalaneni, V. Karuppiah, M. Saleem, L.P. Haynes, R.D. Burgoyne, O. Mayans, et al., Neuronal Calcium Sensor-1 binds the D2 dopamine receptor and G-protein coupled receptor kinase 1 (GRK1) peptides using different modes of interactions, J. Biol. Chem. 1 (2015) jbc–M114.

[14] J.B.. H.K.B.. S.T.. H.I.G.. H.N.. T.J. Ames, Structure and calcium-binding properties of Frq1, a novel calcium sensor in the yeast Saccharomyces cerevisiae., Biochemistry. 39 (2000) 12149–12161.

[15] T. Sippy, A. Cruz-Martín, A. Jeromin, F.E. Schweizer, Acute changes in short-term plasticity at synapses with elevated levels of neuronal calcium sensor-1., Nat. Neurosci. 6 (2003) 1031–1038.

[16] M. Gomez, E. De Castro, E. Guarin, H. Sasakura, A. Kuhara, I. Mori, et al., Ca2+ signaling via the neuronal calcium sensor-1 regulates associative learning and memory in C. elegans, Neuron. 30 (2001) 241–248.

[17] O. Pongs, J. Lindemeier, X.R. Zhu, T. Theil, D. Engelkamp, I. Krah-Jentgens, et al., Frequenin—A novel calcium-binding protein that modulates synaptic efficacy in the drosophila nervous system, Neuron. 11 (1993) 15–28.

[18] B.W. McFerran, M.E. Graham, R.D. Burgoyne, Neuronal Ca2+ sensor 1, the mammalian homologue of frequenin, is expressed in chromaffin and PC12 cells and regulates neurosecretion from dense-core granules, J. Biol. Chem. 273 (1998) 22768–22772.

[19] L.P. Haynes, D.J. Fitzgerald, B. Wareing, D.W. O'Callaghan, A. Morgan, R.D. Burgoyne, Analysis of the interacting partners of the neuronal calcium-binding proteins L-CaBP1, hippocalcin, NCS-1 and neurocalcin delta., Proteomics. 6 (2006) 1822–32.

[20] C. Schlecker, W. Boehmerle, A. Jeromin, B. DeGray, A. Varshney, Y. Sharma, et al., Neuronal calcium sensor-1 enhancement of InsP3 receptor activity is inhibited by therapeutic levels of lithium, J. Clin. Invest. 116 (2006) 1668.

[21] N. Bahi, G. Friocourt, A. Carrié, M.E. Graham, J.L. Weiss, P. Chafey, et al., IL1 receptor accessory protein like, a protein involved in X-linked mental retardation, interacts with Neuronal Calcium Sensor-1 and regulates exocytosis, Hum. Mol. Genet. 12 (2003) 1415–1425.

[22] L.P. Haynes, M.W. Sherwood, N.J. Dolman, R.D. Burgoyne, Specificity, Promiscuity and Localization of ARF Protein Interactions with NCS-1 and Phosphatidylinositol-4 Kinase-IIIβ, Traffic. 8 (2007) 1080–1092.

[23] J. de Barry, A. Janoshazi, J.L. Dupont, O. Procksch, S. Chasserot-Golaz, A. Jeromin, et al., Functional implication of neuronal calcium sensor-1 and phosphoinositol 4-kinase-beta interaction in regulated exocytosis of PC12 cells., J. Biol. Chem. 281 (2006) 18098–111.

[24] H. Hui, D. McHugh, M. Hannan, F. Zeng, S. Xu, S. Khan, et al., Calcium-sensing mechanism in TRPC5 channels contributing to retardation of neurite outgrowth, J. Physiol. 572 (2006) 165–172.

[25] L.P. Haynes, G.M.H. Thomas, R.D. Burgoyne, Interaction of neuronal calcium sensor-1 and ARF1 allows bidirectional control of PI (4) kinase beta and TGN-plasma membrane traffic, J. Biol. Chem. (2004).

[26] K.B. Hendricks, B.Q. Wang, E. a Schnieders, J. Thorner, Yeast homologue of neuronal frequenin is a regulator of phosphatidylinositol-4-OH kinase., Nat. Cell Biol. 1 (1999) 234–241.

[27] N. Kabbani, L. Negyessy, R. Lin, P. Goldman-rakic, R. Levenson, Interaction with Neuronal Calcium Sensor NCS-1 Mediates Desensitization of the D2 Dopamine Receptor, 22 (2002) 8476–8486.

[28] A. Bonci, F.W. Hopf, The dopamine D2 receptor: new surprises from an old friend., Neuron. 47 (2005) 335–8.

[29] E.Y.T. Chien, W. Liu, Q. Zhao, V. Katritch, G.W. Han, M. a Hanson, et al., Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist., Science. 330 (2010) 1091–5.

[30] N. Kabbani, R. Levenson, Antipsychotic-induced alterations in D2 dopamine receptor interacting proteins within the cortex., Neuroreport. 17 (2006) 299–301.

[31] A. Dagher, T.W. Robbins, Personality, addiction, dopamine: insights from Parkinson's disease, Neuron. 61 (2009) 502–510.

[32] C. Bergson, R. Levenson, P.S. Goldman-Rakic, M.S. Lidow, Dopamine receptor-interacting proteins: the Ca(2+) connection in dopamine signaling., Trends Pharmacol. Sci. 24 (2003) 486–92.

[33] Y. Bourne, J. Dannenberg, V. Pollmann, P. Marchot, O. Pongs, Immunocytochemical localization and crystal structure of human frequenin (neuronal calcium sensor 1)., J. Biol. Chem. 276 (2001) 11949–55.

[34] T. Strahl, I.G. Huttner, J.D. Lusin, M. Osawa, D. King, J. Thorner, et al., Structural insights into activation of phosphatidylinositol 4-kinase (Pik1) by yeast frequenin (Frq1)., J. Biol. Chem. 282 (2007) 30949–59.

[35] H. Wang, Y. Yan, Q. Liu, Y. Huang, Y. Shen, L. Chen, et al., Structural basis for modulation of Kv4 K+ channels by auxiliary KChIP subunits, Nat Neurosci. 10 (2007) 32–39.

[36] J.B. Ames, K. Levay, J.N. Wingard, J.D. Lusin, V.Z. Slepak, Structural basis for calcium-induced inhibition of rhodopsin kinase by recoverin, J. Biol. Chem. 281 (2006) 37237–37245.

[37] M. Pioletti, F. Findeisen, G.L. Hura, D.L. Minor, Three-dimensional structure of the KChIP1–Kv4. 3 T1 complex reveals a cross-shaped octamer, Nat. Struct. Mol. Biol. 13 (2006) 987–995.

[38] L.-Y. Lian, S.R. Pandalaneni, P. Patel, H. V McCue, L.P. Haynes, R.D. Burgoyne, Characterisation of the interaction of the C-terminus of the dopamine D2 receptor with neuronal calcium sensor-1., PLoS One. 6 (2011) e27779.

[39] N. Hamasaki-Katagiri, J.B. Ames, Neuronal calcium sensor-1 (Ncs1p) is up-regulated by calcineurin to promote Ca2+ tolerance in fission yeast, J. Biol. Chem. 285 (2010) 4405–4414.

[40] N. Hamasaki-Katagiri, T. Molchanova, K. Takeda, J.B. Ames, Fission Yeast Homolog of Neuronal Calcium Sensor-1 (Ncs1p) Regulates Sporulation and Confers Calcium Tolerance, J. Biol. Chem. 279 (2004) 12744–12754.

[41] T. Strahl, B. Grafelmann, J. Dannenberg, J. Thorner, O. Pongs, Conservation of Regulatory Function in Calcium-binding Proteins HUMAN FREQUENIN (NEURONAL CALCIUM SENSOR-1) ASSOCIATES PRODUCTIVELY WITH YEAST PHOSPHATIDYLINOSITOL 4-KINASE ISOFORM, Pik1, J. Biol. Chem. 278 (2003) 49589–49599.

[42] E. Giralt, M. Peczuh, X. Salvatella, Protein Surface Recognition: Approaches for Drug Discovery, John Wiley & Sons, 2011.

[43] J. a Wells, C.L. McClendon, Reaching for high-hanging fruit in drug discovery at protein-protein interfaces., Nature. 450 (2007) 1001–9.

[44] M.R. Arkin, Y. Tang, J. A Wells, Small-Molecule Inhibitors of Protein-Protein Interactions: Progressing toward the Reality., Chem. Biol. 21 (2014) 1102–1114.

[45] L. Lo Conte, C. Chothia, È. Janin, The Atomic Structure of Protein-Protein Recognition Sites, (1999).

[46] T. Clackson, J.A. Wells, A hot spot of binding energy in a hormone-receptor interface, Science (80-.). 267 (1995) 383–386.

[47] A.A. Bogan, K.S. Thorn, Anatomy of hot spots in protein interfaces, J. Mol. Biol. 280 (1998) 1–9.

[48] W.L. DeLano, Unraveling hot spots in binding interfaces: progress and challenges, Curr. Opin. Struct. Biol. 12 (2002) 14–20.

[49] D.C. Fry, Drug-like inhibitors of protein-protein interactions: a structural examination of effective protein mimicry, Curr. Protein Pept. Sci. 9 (2008) 240–247.

[50] X. Morelli, R. Bourgeas, P. Roche, Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I)., Curr. Opin. Chem. Biol. 15 (2011) 475–81.

[51] L. Laraia, G. McKenzie, D.R. Spring, A.R. Venkitaraman, D.J. Huggins, Overcoming Chemical, Biological, and Computational Challenges in the Development of Inhibitors Targeting Protein-Protein Interactions, Chem. Biol. 22 (2015) 689–703.

[52] A.C. Braisted, J.D. Oslob, W.L. Delano, J. Hyde, R.S. McDowell, N. Waal, et al., Discovery of a potent small molecule IL-2 inhibitor through fragment assembly, J. Am. Chem. Soc. 125 (2003) 3714–3715.

[53] B.C. Raimundo, J.D. Oslob, A.C. Braisted, J. Hyde, R.S. McDowell, M. Randal, et al., Integrating fragment assembly and biophysical methods in the chemical advancement of small-molecule antagonists of IL-2: An approach for inhibiting protein-protein interactions, J. Med. Chem. 47 (2004) 3111–3130.

[54] M.R. Arkin, J. A Wells, Small-molecule inhibitors of protein-protein interactions: progressing towards the dream., Nat. Rev. Drug Discov. 3 (2004) 301–317.

[55] Y. Wang, R. Coulombe, D.R. Cameron, L. Thauvette, M.-J. Massariol, L.M. Amon, et al., Crystal structure of the E2 transactivation domain of human papillomavirus type 11 bound to a protein interaction inhibitor., J. Biol. Chem. 279 (2004) 6976–6985.

[56] C. Yoakim, W.W. Ogilvie, N. Goudreau, J. Naud, B. Haché, J. a. O'Meara, et al., Discovery of the first series of inhibitors of human papillomavirus type 11: Inhibition of the assembly of the E1-E2-Origin DNA complex, Bioorganic Med. Chem. Lett. 13 (2003) 2539–2541.

[57] P.W. White, S. Titolo, K. Brault, L. Thauvette, A. Pelletier, E. Welchner, et al., Inhibition of human papillomavirus DNA replication by small molecule antagonists of the E1-E2 protein interaction., J. Biol. Chem. 278 (2003) 26765–26772.

[58] N. Goudreau, D.R. Cameron, R. Déziel, B. Haché, A. Jakalian, E. Malenfant, et al., Optimization and determination of the absolute configuration of a series of potent inhibitors of human papillomavirus type-11 E1-E2 protein-protein interaction: A combined medicinal chemistry, NMR and computational chemistry approach, Bioorganic Med. Chem. 15 (2007) 2690–2700.

[59] L. Mosyak, Y. Zhang, E. Glasfeld, S. Haney, M. Stahl, J. Seehra, et al., The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography., EMBO J. 19 (2000) 3179–3191.

[60] D.H.H. Tsao, A.G. Sutherland, L.D. Jennings, Y. Li, T.S. Rush, J.C. Alvarez, et al., Discovery of novel inhibitors of the ZipA/FtsZ complex by NMR fragment screening coupled with structure-based design, Bioorganic Med. Chem. 14 (2006) 7953–7961.

[61] L.D. Jennings, K.W. Foreman, T.S. Rush, D.H.H. Tsao, L. Mosyak, S.L. Kincaid, et al., Combinatorial synthesis of substituted 3-(2-indolyl)piperidines and 2-phenyl indoles as inhibitors of ZipA-FtsZ interaction, Bioorganic Med. Chem. 12 (2004) 5115–5131.

[62] P.H. Carter, P.A. Scherle, J.A. Muckelbauer, M.E. Voss, R.-Q. Liu, L.A. Thompson, et al., Photochemically enhanced binding of small molecules to the tumor necrosis factor receptor-1 inhibits the binding of TNF-α, Proc. Natl. Acad. Sci. 98 (2001) 11879–11884.

[63] M.M. He, A.S. Smith, J.D. Oslob, W.M. Flanagan, A.C. Braisted, A. Whitty, et al., Small-Molecule Inhibition of TNF- a, (n.d.).

[64] a J. Levine, W. Hu, Z. Feng, The P53 pathway: what questions remain to be explored?, Cell Death Differ. 13 (2006) 1027–1036.

[65] P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, a J. Levine, et al., Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain., Science. 274 (1996) 948–953.

[66] W. Tao, A.J. Levine, Nucleocytoplasmic shuttling of oncoprotein Hdm2 is required for Hdm2-mediated degradation of p53, Proc. Natl. Acad. Sci. 96 (1999) 3077–3080.

[67] M.D. Cummings, C. Schubert, D.J. Parks, R.R. Calvo, L. V LaFrance, J. Lattanze, et al., Substituted 1,4-benzodiazepine-2,5-diones as alpha-helix mimetic antagonists of the HDM2-p53 protein-protein interaction., Chem. Biol. Drug Des. 67 (2006) 201–5.

[68] K. Ding, Y. Lu, Z. Nikolovska-coleska, S. Qiu, Y. Ding, W. Gao, et al., Structure-Based Design of Potent Non-Peptide MDM2 Inhibitors, (2005) 10130–10131.

[69] K. Ding, Y. Lu, Z. Nikolovska-Coleska, G. Wang, S. Qiu, S. Shangary, et al., Structure-based design of spiro-oxindoles as potent, specific small-molecule inhibitors of the MDM2-p53 interaction, J. Med. Chem. 49 (2006) 3432–3435.

[70] C. Klein, In Vivo Activation of the p53 Pathway by Small- Molecule Antagonists of MDM2, (2004).

[71] B.L. Grasberger, T. Lu, C. Schubert, D.J. Parks, T.E. Carver, H.K. Koblish, et al., Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells, J. Med. Chem. 48 (2005) 909–912.

[72] Y. Rew, D. Sun, F. Gonzalez-Lopez De Turiso, M.D. Bartberger, H.P. Beck, J. Canon, et al., Structure-based design of novel inhibitors of the MDM2-p53 interaction., J. Med. Chem. 55 (2012) 4936–54.

[73] M.P. Dickens, R. Fitzgerald, P.M. Fischer, Small-molecule inhibitors of MDM2 as new anticancer therapeutics, Semin. Cancer Biol. 20 (2010) 10–18.

[74] M. Sattler, Structure of Bcl-xL-Bak Peptide Complex: Recognition Between Regulators of Apoptosis, Science (80-.). 275 (1997) 983–986.

[75] J.M. Rodriguez, N.T. Ross, W.P. Katt, D. Dhar, G.-I. Lee, A.D. Hamilton, Structure and function of benzoylurea-derived alpha-helix mimetics targeting the Bcl-x(L)/Bak binding interface., ChemMedChem. 4 (2009) 649–56.

[76] H. Yin, G. Lee, K.A. Sedey, O. Kutzki, H.S. Park, B.P. Orner, et al., Terphenyl-Based Bak BH3 r -Helical Proteomimetics as Low-Molecular-Weight Antagonists of Bcl-x L, (2005) 10191–10196.

[77] H. Yin, G. Lee, K.A. Sedey, J.M. Rodriguez, H. Wang, S.M. Sebti, et al., Terephthalamide Derivatives as Mimetics of Helical Peptides : Disruption of the Bcl-x L / Bak Interaction, (2005) 5463–5468.

[78] J.M. Rodriguez, A.D. Hamilton, Intramolecular hydrogen bonding allows simple enaminones to structurally mimic the i , i + 4 , and i + 7 residues of an a -helix, 47 (2006) 7443–7446.

[79] C. Reynès, H. Host, A.-C. Camproux, G. Laconde, F. Leroux, A. Mazars, et al., Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods, PLoS Comput. Biol. 6 (2010) e1000695.

[80] G.M. Keserü, G.M. Makara, The influence of lead discovery strategies on the properties of drug candidates, Nat. Rev. Drug Discov. 8 (2009) 203–212.

[81] R.A. Goodnow, Hit and lead identification: Integrated technology-based approaches, Drug Discov. Today Technol. 3 (2006) 367–375.

[82] S. Kalyaanamoorthy, Y.-P.P. Chen, Structure-based drug design to augment hit discovery., Drug Discov. Today. 16 (2011) 831–9.

[83] S.-Y. Yang, Pharmacophore modeling and applications in drug discovery: challenges and recent advances, Drug Discov. Today. 15 (2010) 444–450.

[84] A. Varnek, A. Tropsha, Chemoinformatics approaches to virtual screening, Royal Society of Chemistry, 2008.

[85] G. Wolber, T. Seidel, F. Bendix, T. Langer, Molecule-pharmacophore superpositioning and pattern matching in computational drug design, Drug Discov. Today. 13 (2008) 23–29.

[86] K. Poptodorov, T. Luu, R.D. Hoffmann, Pharmacophore model generation software tools, Methods Princ. Med. Chem. 32 (2006) 17.

[87] C. Barillari, G. Marcou, D. Rognan, Hot-spots-guided receptor-based pharmacophores (HS-Pharm): a knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores, J. Chem. Inf. Model. 48 (2008) 1396–1410.

[88] K.-H. Kim, N.D. Kim, B.-L. Seong, Pharmacophore-based virtual screening: a review of recent applications, Expert Opin. Drug Discov. 5 (2010) 205–222.

[89] C.G. Wermuth, Pharmacophores: historical perspective and viewpoint from a medicinal chemist, Methods Princ. Med. Chem. 32 (2006) 3.

[90] T. Hou, J. Wang, W. Zhang, W. Wang, X. Xu, Recent advances in computational prediction of drug absorption and permeability in drug discovery, Curr. Med. Chem. 13 (2006) 2653–2667.

[91] O. Dror, A. Shulman-Peleg, R. Nussinov, H.J. Wolfson, Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design, Curr. Med. Chem. 11 (2004) 71–90.

[92] D.R. Koes, C.J. Camacho, ZINCPharmer: pharmacophore search of the ZINC database, Nucleic Acids Res. . 40 (2012) W409–W414.

[93] D.R. Koes, C.J. Camacho, Pharmer: efficient and exact pharmacophore search, J. Chem. Inf. Model. 51 (2011) 1307–1314.

[94] J.S. Mason, D.L. Cheney, Library design and virtual screening using multiple 4-point pharmacophore fingerprints, in: Pac. Symp. Biocomput, World Scientific, 2000: pp. 576–587.

[95] D.F. Veber, S.R. Johnson, H.-Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates., J. Med. Chem. 45 (2002) 2615–2623.

[96] R.D. Taylor, P.J. Jewsbury, J.W. Essex, A review of protein-small molecule docking methods, J. Comput. Aided. Mol. Des. 16 (2002) 151–166.

[97] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[98] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, D.M. Ferguson, et al., AMBER 5.0, Univ. California, San Fr. (1997).

[99] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray, R.D. Taylor, Improved Protein – Ligand Docking Using GOLD, 623 (2003) 609–623.

[100] T.N. Hart, R.J. Read, A multiple-start Monte Carlo docking method, Proteins Struct. Funct. Bioinforma. 13 (1992) 206–222.

[101] D.S. Goodsell, G.M. Morris, A.J. Olson, Automated docking of flexible ligands: applications of AutoDock, J. Mol. Recognit. 9 (1996) 1–5.

[102] D.S. Goodsell, A.J. Olson, Automated docking of substrates to proteins by simulated annealing, Proteins Struct. Funct. Bioinforma. 8 (1990) 195–202.

[103] J.M. Blaney, J.S. Dixon, DockIt, version 1.0; Metaphorics, LLC: Mission Viejo, CA,

[104] A.R. Leach, A.S. Smellie, A combined model-building and distance-geometry approach to automated conformational analysis and search, J. Chem. Inf. Comput. Sci. 32 (1992) 379–385.

[105] B. Kramer, M. Rarey, T. Lengauer, Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking, Proteins Struct. Funct. Bioinforma. 37 (1999) 228–241.

[106] M. Rarey, B. Kramer, T. Lengauer, G. Klebe, A fast flexible docking method using an incremental construction algorithm, J. Mol. Biol. 261 (1996) 470–489.

[107] C.A. Baxter, C.W. Murray, D.E. Clark, D.R. Westhead, M.D. Eldridge, Flexible docking using Tabu search and an empirical estimate of binding affinity, Proteins Struct. Funct. Bioinforma. 33 (1998) 367–382.

[108] E. Perola, K. Xu, T.M. Kollmeyer, S.H. Kaufmann, F.G. Prendergast, Y.-P. Pang, Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads, J. Med. Chem. 43 (2000) 401–408.

[109] G. Jones, P. Willett, R.C. Glen, Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation, J. Mol. Biol. 245 (1995) 43–53.

[110] J.W.M. Nissink, C. Murray, M. Hartshorn, M.L. Verdonk, J.C. Cole, R. Taylor, A new test set for validating predictions of protein-ligand interaction, Proteins Struct. Funct. Genet. 49 (2002) 457–471.

[111] W.T.M. Mooij, M.L. Verdonk, General and targeted statistical potentials for protein-ligand interactions., Proteins. 61 (2005) 272–87.

[112] G.M. Verkhivker, Computational analysis of ligand binding dynamics at the intermolecular hot spots with the aid of simulated tempering and binding free energy calculations., J. Mol. Graph. Model. 22 (2004) 335–48.

[113] D.J. Livingstone, A. Davis, D.E. Thurston, Drug design strategies: Quantitative approaches, Royal Society of Chemistry, 2011.

[114] M.A. Navia, P.R. Chaturvedi, Design principles for orally bioavailable drugs, Drug Discov. Today. 1 (1996) 179–189.

[115] R. Hirschmann, Peptide related research: A means to further biological and chemical understanding, in: Pept. Symp., ESCOM SCIENCE PUBLISHERS, 1996: pp. 3–20.

[116] C. a. Lipinski, Lead- and drug-like compounds: The rule-of-five revolution, Drug Discov. Today Technol. 1 (2004) 337–341.

[117] M.M. Hann, Molecular obesity, potency and other addictions in drug discovery, Medchemcomm. 2 (2011) 349–355.

[118] M.P. Gleeson, Generation of a set of simple, interpretable ADMET rules of thumb, J. Med. Chem. 51 (2008) 817–834.

[119] J.D. Hughes, J. Blagg, D.A. Price, S. Bailey, G.A. DeCrescenzo, R. V Devraj, et al., Physiochemical drug properties associated with in vivo toxicological outcomes, Bioorg. Med. Chem. Lett. 18 (2008) 4872–4875.

[120] P.D. Leeson, B. Springthorpe, The influence of drug-like concepts on decision-making in medicinal chemistry, Nat. Rev. Drug Discov. 6 (2007) 881–890.

[121] T.J. Ritchie, S.J.F. Macdonald, The impact of aromatic ring count on compound developability–are too many aromatic rings a liability in drug design?, Drug Discov. Today. 14 (2009) 1011–1020.

[122] A.L. Hopkins, C.R. Groom, A. Alex, Ligand efficiency: a useful metric for lead selection, Drug Discov. Today. 9 (2004) 430–431.

[123] O. Nicolotti, I. Giangreco, A. Introcaso, F. Leonetti, A. Stefanachi, A. Carotti, Strategies of multi-objective optimization in drug discovery and development, Expert Opin. Drug Discov. 6 (2011) 871–884.

[124] M.R. Berthold, N. Cebron, F. Dill, T.R. Gabriel, T. Kötter, T. Meinl, et al., KNIME: The Konstanz information miner, Springer, 2008.

[125] A.S. Inc, Pipeline Pilot Student Edn, (n.d.).

[126] W.A. Warr, Scientific workflow systems: Pipeline Pilot and KNIME, J. Comput. Aided. Mol. Des. 26 (2012) 801–804.

[127] Eclipse., No Title, (n.d.).

[128] K.D. Stigers, M.J. Soth, J.S. Nowick, Designed molecules that fold to mimic protein secondary structures, Curr. Opin. Chem. Biol. 3 (1999) 714–723.

[129] S. Fletcher, A.D. Hamilton, Protein surface recognition and proteomimetics: mimics of protein surface structure and function, Curr. Opin. Chem. Biol. 9 (2005) 632–638.

[130] M. Jesus Perez de Vega, M. Martín-Martínez, R. González-Muñiz, Modulation of protein-protein interactions by stabilizing/mimicking protein secondary structure elements, Curr. Top. Med. Chem. 7 (2007) 33–62.

[131] K. Burgess, Solid-phase syntheses of β-turn analogues to mimic or disrupt protein-protein interactions, Acc. Chem. Res. 34 (2001) 826–835.

[132] J.A. Kritzer, O.M. Stephens, D.A. Guarracino, S.K. Reznik, A. Schepartz, β-peptides as inhibitors of protein–protein interactions, Bioorg. Med. Chem. 13 (2005) 11–16.

[133] L. Zhao, J. Chmielewski, Inhibiting protein–protein interactions using designed molecules, Curr. Opin. Struct. Biol. 15 (2005) 31–34.

[134] B.A. Bunin, J.A. Ellman, A general and expedient method for the solid-phase synthesis of 1, 4-benzodiazepine derivatives, J. Am. Chem. Soc. 114 (1992) 10997–10998.

[135] C. Gil, S. Bräse, Efficient Solid-Phase Synthesis of Highly Functionalized 1, 4-Benzodiazepin-5-one Derivatives and Related Compounds by Intramolecular Aza–Wittig Reactions, Chem. Eur. J. 11 (2005) 2680–2688.

[136] D.J. Parks, L. V LaFrance, R.R. Calvo, K.L. Milkiewicz, V. Gupta, J. Lattanze, et al., 1, 4-Benzodiazepine-2, 5-diones as small molecule antagonists of the HDM2–p53 interaction: discovery and SAR, Bioorg. Med. Chem. Lett. 15 (2005) 765–770.

[137] B.P. Orner, J.T. Ernst, A.D. Hamilton, Toward proteomimetics: terphenyl derivatives as structural and functional mimics of extended regions of an α-helix, J. Am. Chem. Soc. 123 (2001) 5382–5383.

[138] J.T. Ernst, J. Becerril, H.S. Park, H. Yin, A.D. Hamilton, Design and Application of an α-Helix-Mimetic Scaffold Based on an Oligoamide-Foldamer Strategy: Antagonism of the Bak BH3/Bcl-xL Complex, Angew. Chemie. 115 (2003) 553–557.

[139] I.C. Kim, A.D. Hamilton, Diphenylindane-based proteomimetics reproduce the projection of the i, i+ 3, i+ 4, and i+ 7 residues on an α-helix, Org. Lett. 8 (2006) 1751–1754.

[140] W.P. Nolan, G.S. Ratcliffe, D.C. Rees, The synthesis of 1, 6-disubstituted indanes which mimic the orientation of amino acid side-chains in a protein alpha-helix motif., Tetrahedron Lett. 33 (1992) 6879–6882.

[141] I.R. Hardcastle, S.U. Ahmed, H. Atkins, G. Farnie, B.T. Golding, R.J. Griffin, et al., Small-molecule inhibitors of the MDM2-p53 protein-protein interaction based on an isoindolinone scaffold, J. Med. Chem. 49 (2006) 6209–6221.

[142] J.M. Davis, A. Truong, A.D. Hamilton, Synthesis of a 2, 3'; 6', 3"-Terpyridine Scaffold as an α-Helix Mimetic, Org. Lett. 7 (2005) 5405–5408.

[143] H. Oguri, S. Tanabe, A. Oomura, M. Umetsu, M. Hirama, Synthesis and evaluation of α-helix mimetics based on a trans-fused polycyclic ether: sequence-selective binding to aspartate pairs in α-helical peptides, Tetrahedron Lett. 47 (2006) 5801–5805.

[144] J.-M. Ahn, S.-Y. Han, Facile synthesis of benzamides to mimic an α-helix, Tetrahedron Lett. 48 (2007) 3543–3547.

[145] W. Antuch, S. Menon, Q.-Z. Chen, Y. Lu, S. Sakamuri, B. Beck, et al., Design and modular parallel synthesis of a MCR derived α-helix mimetic protein–protein interaction inhibitor scaffold, Bioorg. Med. Chem. Lett. 16 (2006) 1740–1743.

[146] D.L. Boger, J. Desharnais, K. Capps, Solution-Phase Combinatorial Libraries: Modulating Cellular Signaling by Targeting Protein–Protein or Protein–DNA Interactions, Angew. Chemie Int. Ed. 42 (2003) 4138–4176.

[147] D.L. Boger, J.K. Lee, J. Goldberg, Q. Jin, Two comparisons of the performance of positional scanning and deletion synthesis for the identification of active constituents in mixture combinatorial libraries, J. Org. Chem. 65 (2000) 1467–1474.

[148] S. Cheng, C.M. Tarby, D.D. Comer, J.P. Williams, L.H. Caporale, P.L. Myers, et al., A solution-phase strategy for the synthesis of chemical libraries containing small organic molecules: a universal and dipeptide mimetic template, Bioorg. Med. Chem. 4 (1996) 727–737.

[149] T. Berg, S.B. Cohen, J. Desharnais, C. Sonderegger, D.J. Maslyar, J. Goldberg, et al., Small-molecule antagonists of Myc/Max dimerization inhibit Myc-induced transformation of chicken embryo fibroblasts, Proc. Natl. Acad. Sci. 99 (2002) 3830–3835.

[150] M. Eguchi, M.S. Lee, H. Nakanishi, M. Stasiak, S. Lovell, M. Kahn, Solid-phase synthesis and structural analysis of bicyclic β-turn mimetics incorporating functionality at the i to i+ 3 positions, J. Am. Chem. Soc. 121 (1999) 12204–12205.

[151] A. Golebiowski, S.R. Klopfenstein, J.J. Chen, X. Shao, Solid supported high-throughput organic synthesis of peptide β-turn mimetics via tandem Petasis reaction/diketopiperazine formation, Tetrahedron Lett. 41 (2000) 4841–4844.

[152] A. Golebiowski, S.R. Klopfenstein, X. Shao, J.J. Chen, A.-O. Colson, A.L. Grieb, et al., Solid-supported synthesis of a peptide β-turn mimetic, Org. Lett. 2 (2000) 2615–2617.

[153] K.H. Emami, C. Nguyen, H. Ma, D.H. Kim, K.W. Jeong, M. Eguchi, et al., A small molecule inhibitor of β-catenin/cyclic AMP response element-binding protein transcription, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 12682–12687.

[154] M. Eguchi, C. Nguyen, S.C. Lee, M. Kahn, ICG-001, A Novel Small Molecule Regulator of TCF/β-Catenin Masakatsu Transcription, Med. Chem. (Los. Angeles). 1 (2005) 467–472.

[155] D.J. Newman, G.M. Cragg, K.M. Snader, Natural products as sources of new drugs over the period 1981-2002, J. Nat. Prod. 66 (2003) 1022–1037.

[156] R. Breinbauer, I.R. Vetter, H. Waldmann, From protein domains to drug candidates—natural products as guiding principles in the design and synthesis of compound libraries, Angew. Chemie Int. Ed. 41 (2002) 2878–2890.

[157] S. Shang, D.S. Tan, Advancing chemistry and biology through diversity-oriented synthesis of natural product-like libraries, Curr. Opin. Chem. Biol. 9 (2005) 248–258.

[158] A. Reayi, P. Arya, Natural product-like chemical space: search for chemical dissectors of macromolecular interactions, Curr. Opin. Chem. Biol. 9 (2005) 240–247.

[159] A.D. Patil, A.J. Freyer, P.B. Taylor, B. Carté, G. Zuber, R.K. Johnson, et al., Batzelladines FI, novel alkaloids from the sponge Batzella sp.: Inducers of p56lck-CD4 dissociation, J. Org. Chem. 62 (1997) 1814–1819.

[160] A.D. Patil, A.J. Freyer, P. Offen, M.F. Bean, R.K. Johnson, Three new tricyclic guanidine alkaloids from the sponge Batzella sp., J. Nat. Prod. 60 (1997) 704–707.

[161] A.D. Patil, N.V. Kumar, W.C. Kokke, M.F. Bean, A.J. Freyer, C. De Brosse, et al., Novel alkaloids from the sponge Batzella sp.: inhibitors of HIV gp120-human CD4 binding, J. Org. Chem. 60 (1995) 1182–1188.

[162] A. Olszewski, G.A. Weiss, Library versus library recognition and inhibition of the HIV-1 Nef allelome, J. Am. Chem. Soc. 127 (2005) 12178–12179.

[163] B.B. Snider, M. V Busuyek, Revision of the stereochemistry of batzelladine F. Approaches to the tricyclic hydroxyguanidine moiety of batzelladines G, H, and I, J. Nat. Prod. 62 (1999) 1707–1711.

[164] C.A. Bewley, S. Ray, F. Cohen, S.K. Collins, L.E. Overman, Inhibition of HIV-1 envelope-mediated fusion by synthetic batzelladine analogues, J. Nat. Prod. 67 (2004) 1319–1324.

[165] A. Olszewski, K. Sato, Z.D. Aron, F. Cohen, A. Harris, B.R. McDougall, et al., Guanidine alkaloid analogs as inhibitors of HIV-1 Nef interactions with p53, actin, and p56lck, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 14079–14084.

[166] F. Cohen, L.E. Overman, Enantioselective total synthesis of batzelladine F and definition of its structure, J. Am. Chem. Soc. 128 (2006) 2604–2608.

[167] R. Bai, G.F. Taylor, Z.A. Cichacz, C.L. Herald, J.A. Kepler, G.R. Pettit, et al., The spongistatins, potently cytotoxic inhibitors of tubulin polymerization, bind in a distinct region of the vinca domain, Biochemistry. 34 (1995) 9714–9721.

[168] R.F. Luduena, M.C. Roach, V. Prasad, G.R. Pettit, Z.A. Cichacz, C.L. Herald, Interaction of three sponge-derived macrocyclic lactone polyethers (spongistatin 3, halistatins 1 and 2) with tubulin, Drug Dev. Res. 35 (1995) 40–48.

[169] R. Bai, Z.A. Cichacz, C.L. Herald, G.R. Pettit, E. Hamel, Spongistatin 1, a highly cytotoxic, sponge-derived, marine natural product that inhibits mitosis, microtubule assembly, and the binding of vinblastine to tubulin., Mol. Pharmacol. 44 (1993) 757–766.

[170] X. Fan, G.R. Flentke, D.H. Rich, Inhibition of HIV-1 protease by a subunit of didemnaketal A, J. Am. Chem. Soc. 120 (1998) 8893–8894.

[171] A. V Statsuk, R. Bai, J.L. Baryza, V.A. Verma, E. Hamel, P.A. Wender, et al., Actin is the primary cellular receptor of bistramide A, Nat. Chem. Biol. 1 (2005) 383–388.

[172] A. V Statsuk, D. Liu, S.A. Kozmin, Synthesis of bistramide A, J. Am. Chem. Soc. 126 (2004) 9546–9547.

[173] B.A. Kulkarni, G.P. Roth, E. Lobkovsky, J.A. Porco, Combinatorial synthesis of natural product-like molecules using a first-generation spiroketal scaffold, J. Comb. Chem. 4 (2002) 56–72.

[174] G. Zinzalla, L.-G. Milroy, S. V Ley, Chemical variation of natural product-like scaffolds: design and synthesis of spiroketal derivatives, Org. Biomol. Chem. 4 (2006) 1977–2002.

[175] S.B. Moilanen, J.S. Potuzak, D.S. Tan, Stereocontrolled Synthesis of Spiroketals via Ti (O i-Pr) 4-Mediated Kinetic Spirocyclization of Glycal Epoxides with Retention of Configuration, J. Am. Chem. Soc. 128 (2006) 1792–1793.

[176] J.S. Potuzak, S.B. Moilanen, D.S. Tan, Stereocontrolled synthesis of spiroketals via a remarkable methanol-induced kinetic spirocyclization reaction, J. Am. Chem. Soc. 127 (2005) 13796–13797.

[177] C. V Galliford, K.A. Scheidt, Pyrrolidinyl-Spirooxindole Natural Products as Inspirations for the Development of Potential Therapeutic Agents, Angew. Chemie Int. Ed. 46 (2007) 8748–8758.

[178] P.R. Sebahar, H. Osada, T. Usui, R.M. Williams, Asymmetric, stereocontrolled total synthesis of (+) and (−)-spirotryprostatin B via a diastereoselective azomethine ylide [1, 3]-dipolar cycloaddition reaction, Tetrahedron. 58 (2002) 6311–6322.

[179] P.R. Sebahar, R.M. Williams, The synthesis of spirooxindole pyrrolidines via an asymmetric azomethine ylide [1, 3]-dipolar cycloaddition reaction, Heterocycles. 58 (2002) 563–575.

[180] P.R. Sebahar, R.M. Williams, The asymmetric total synthesis of (+)-and (-)-spirotryprostatin B, J. Am. Chem. Soc. 122 (2000) 5666–5667.

[181] M.M.-C. Lo, C.S. Neumann, S. Nagayama, E.O. Perlstein, S.L. Schreiber, A library of spirooxindoles based on a stereoselective three-component coupling reaction, J. Am. Chem. Soc. 126 (2004) 16077–16086.

[182] C. Chen, X. Li, S.L. Schreiber, Catalytic asymmetric [3+ 2] cycloaddition of azomethine ylides. Development of a versatile stepwise, three-component reaction for diversity-oriented synthesis, J. Am. Chem. Soc. 125 (2003) 10174–10175.

[183] R.J. Spandl, M. Díaz-Gavilán, K.M.G. O'Connell, G.L. Thomas, D.R. Spring, Diversity-oriented synthesis, Chem. Rec. 8 (2008) 129–142.

[184] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25.

[185] M. Congreve, R. Carr, C. Murray, H. Jhoti, A "rule of three"for fragment-based lead discovery?, Drug Discov. Today. 8 (2003) 876–877.

[186] M.M. Hann, T.I. Oprea, Pursuing the leadlikeness concept in pharmaceutical research, (n.d.).

[187] S.L. Schreiber, Target-oriented and diversity-oriented organic synthesis in drug discovery, Science (80-.). 287 (2000) 1964–1969.

[188] P. Arya, R. Joseph, Z. Gan, B. Rakic, Exploring new chemical space by stereocontrolled diversity-oriented synthesis, Chem. Biol. 12 (2005) 163–180.

[189] M.D. Burke, S.L. Schreiber, A Planning Strategy for Diversity-Oriented Synthesis, Angew. Chemie Int. Ed. 43 (2004) 46–58.

[190] D.S. Tan, Diversity-oriented synthesis: exploring the intersections between chemistry and biology, Nat. Chem. Biol. 1 (2005) 74–84.

[191] G.L. Thomas, E.E. Wyatt, D.R. Spring, Enriching chemical space with diversity-oriented synthesis, Curr Opin Drug Disc Dev. (2006) 700–712.

[192] D.L. Boger, W. Chai, R.S. Ozer, C.-M. Andersson, Solution-phase combinatorial synthesis via the olefin metathesis reaction, Bioorg. Med. Chem. Lett. 7 (1997) 463–468.

[193] D.L. Boger, W. Chai, Solution-phase combinatorial synthesis: convergent multiplication of diversity via the olefin metathesis reaction, Tetrahedron. 54 (1998) 3955–3970.

[194] D.L. Boger, R.S. Ozer, C.-M. Andersson, Generation of targeted C 2-symmetrical compound libraries by solution-phase combinatorial chemistry, Bioorg. Med. Chem. Lett. 7 (1997) 1903–1908.

[195] D.L. Boger, J. Goldberg, W. Jiang, W. Chai, P. Ducray, J.K. Lee, et al., Higher order iminodiacetic acid libraries for probing protein–protein interactions, Bioorg. Med. Chem. 6 (1998) 1347–1378.

[196] Y. Xu, J. Shi, N. Yamamoto, J.A. Moss, P.K. Vogt, K.D. Janda, A credit-card library approach for disrupting protein–protein interactions, Bioorg. Med. Chem. 14 (2006) 2660–2673.

[197] Y. Xu, H. Lu, J.P. Kennedy, X. Yan, L.A. McAllister, N. Yamamoto, et al., Evaluation of “credit card” libraries for inhibition of HIV-1 gp41 fusogenic core formation, J. Comb. Chem. 8 (2006) 531–539.

[198] S.M. Sternson, J.B. Louca, J.C. Wong, S.L. Schreiber, Split - Pool Synthesis of 1 , 3-Dioxanes Leading to Arrayed Stock Solutions of Single Compounds Sufficient for Multiple Phenotypic and Protein-Binding Assays, (2001) 1740–1747.

[199] R.A. Stavenger, S.L. Schreiber, Asymmetric Catalysis in Diversity-Oriented Organic Synthesis: Enantioselective Synthesis of 4320 Encoded and Spatially Segregated Dihydropyrancarboxamides, Angew. Chemie Int. Ed. 40 (2001) 3417–3421.

[200] K. Bedjeguelal, H. Bienayme, A. Dumoulin, S. Poigny, P. Schmitt, E. Tam, Discovery of protein–protein binding disruptors using multi-component condensations small molecules, Bioorg. Med. Chem. Lett. 16 (2006) 3998–4001.

[201] M. Billeter, G. Wagner, K. Wüthrich, Solution NMR structure determination of proteins revisited, J. Biomol. NMR. 42 (2008) 155–158.

[202] C.E. Housecroft, Chemistry: An Introduction to Organic, Inorganic and Physical Chemistry.–3rd edition/Catherine E. Housecroft, Edwin C. Constable, Edinbg. Gate Pearson Educ. Ltd. 1285 (2006) 530–574.

[203] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, Protein NMR spectroscopy: principles and practice, Academic Press, 1995.

[204] K. Wuthrich, NMR of proteins and nucleic acids, Wiley, 1986.

[205] M. Bailyn, A survey of thermodynamics, AIP, 1994.

[206] S.M. Sarge, CO, (2003) 9–18.

[207] J.J. Christensen, R.M. Izatt, L.D. Hansen, J.A. Partridge, Entropy Titration. A Calorimetric Method for the Determination of ΔG, ΔH, and ΔS from a Single Thermometric Titration1a, b, J. Phys. Chem. 70 (1966) 2003–2010.

[208] J.J. Christensen, D.P. Wrathall, R.M. Izatt, Calorimetric Determination of log K, ΔH°, and ΔS° from Thermometric Titration Data., Anal. Chem. 40 (1968) 175–181.

[209] T. Wiseman, S. Williston, J.F. Brandts, L.-N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter, Anal. Biochem. 179 (1989) 131–137.

[210] S. Milev, Isothermal titration calorimetry : Principles and experimental design Agenda Overview of Isothermal Titration Calorimetry ITC experimental design Data analysis Troubleshooting, GE Healthc. (2013) 1–87.

[211] E. Freire, Isothermal Titration Calorimetry and Drug Design, Appl. Note. ©MicroCal, (2006) 1–4.

[212] G.E. Healthcare, L. Sciences, User Manual MicroCal iTC200, (2012).

[213] M.W. Freyer, E. a. Lewis, Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions, Methods Cell Biol. 84 (2008) 79–113.

[214] A. Alderton, P. Davies, K. Illman, D.R. Brown, Ancient conserved domain protein-1 binds copper and modifies its retention in cells, J. Neurochem. 103 (2007) 312–321.

[215] H. Ohtaka, A. Velázquez-Campoy, D. Xie, E. Freire, Overcoming drug resistance in HIV-1 chemotherapy: The binding thermodynamics of Amprenavir and TMC-126 to wild-type and drug-resistant mutants of the HIV-1 protease, Protein Sci. 11 (2002) 1908–1916.

[216] J.R. Lakowicz, Principles of fluorescence spectroscopy, Springer Science & Business Media, 2007.

[217] F. Polarization, An Introduction to Fluorescence Polarization, (1926).

[218] A.M. Rossi, C.W. Taylor, Analysis of protein-ligand interactions by fluorescence polarization, Nat. Protoc. 6 (2011) 365–387.

[219] J. Lakowicz, ed., Fluorescence Anisotropy, in: Princ. Fluoresc. Spectrosc. SE - 10, Springer US, 2006: pp. 353–382.

[220] M.P. Woll, D. a De Cotiis, M.C. Bewley, D.M. Tacelosky, R. Levenson, J.M. Flanagan, Interaction between the D2 dopamine receptor and neuronal calcium sensor-1 analyzed by fluorescence anisotropy., Biochemistry. 50 (2011) 8780–91.

[221] R.B. Sekar, A. Periasamy, Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations, J. Cell Biol. 160 (2003) 629–633.

[222] C.P. Pan, P.L. Muiño, M.D. Barkley, P.R. Callis, Correlation of tryptophan fluorescence spectral shifts and lifetimes arising directly from heterogeneous environment, J. Phys. Chem. B. 115 (2011) 3245–3253.

[223] S.T. Olson, J.D. Shore, Binding of high affinity heparin to antithrombin III. Characterization of the protein fluorescence enhancement., J. Biol. Chem. 256 (1981) 11065–11072.

[224] Schrodinger LLC, The PyMOL Molecular Graphics System, Version 1.3r1, (2010).

[225] Y. Shao, L.F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S.T. Brown, et al., Spartan'08, Wavefunction, Inc. Irvine, CA, Phys. Chem. Chem. Phys. 8 (2006) 3172–3191.

[226] S. Lim, A.M. Dizhoor, J.B. Ames, Structural diversity of neuronal calcium sensor proteins and insights for activation of retinal guanylyl cyclase by GCAP1., Front. Mol. Neurosci. 7 (2014) 19.

[227] G. Klebe, Recent developments in structure-based drug design., J. Mol. Med. (Berl). 78 (2000) 269–281.

[228] A.C. Anderson, The Process of Structure-Based Drug Design, Chem. Biol. 10 (2003) 787–797.

[229] G. Thomas, Medicinal chemistry: an introduction, John Wiley & Sons, 2011.

[230] B.K. Shoichet, Virtual screening of chemical libraries, Nature. 432 (2004) 862–865.

[231] J.A. Wells, C.L. McClendon, Reaching for high-hanging fruit in drug discovery at protein–protein interfaces, Nature. 450 (2007) 1001–1009.

[232] P. Ehrlich, Über den jetzigen Stand der Chemotherapie, Berichte Der Dtsch. Chem. Gesellschaft. 42 (1909) 17–47.

[233] T.. H.I.G.. L.J.D.. O.M.. K.D.. T.J.. A.J.B. Strahl, Structural insights into activation of phosphatidylinositol 4-kinase (Pik1) by yeast frequenin (Frq1)., J.Biol.Chem. 282 (2007) 30949–30959.

[234] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 64 (2012) 4–17.

[235] J.U. Peters, P. Schnider, P. Mattei, M. Kansy, Pharmacological promiscuity: Dependence on compound properties and target specificity in a set of recent roche compounds, ChemMedChem. 4 (2009) 680–686.

[236] V.N. Maiorov, G.M. Crippen, Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins, J. Mol. Biol. 235 (1994) 625–634.

[237] J. Peters, P. Schnider, P. Mattei, M. Kansy, Pharmacological promiscuity: dependence on compound properties and target specificity in a set of recent Roche compounds, ChemMedChem. 4 (2009) 680–686.

[238] D.B. Kitchen, H. Decornez, J.R. Furr, J. Bajorath, Docking and scoring in virtual screening for drug discovery: methods and applications, Nat. Rev. Drug Discov. 3 (2004) 935–949.

[239] C.H. Reynolds, B. a Tounge, S.D. Bembenek, Ligand binding efficiency: trends, physical basis, and implications., J. Med. Chem. 51 (2008) 2432–8.

[240] D.D.T. Vol, N. May, update discussion forum Ligand efficiency : a useful Promiscuity : what protects us , perplexes us, Drug Discov. Today. 9 (2004) 430–431.

[241] J.C. Baber, W.A. Shirley, Y. Gao, M. Feher, The use of consensus scoring in ligand-based virtual screening, J. Chem. Inf. Model. 46 (2006) 277–288.

[242] M. Feher, Consensus scoring for protein-ligand interactions, Drug Discov. Today. 11 (2006) 421–428.

[243] C. a Nicolaou, N. Brown, C.S. Pattichis, Molecular optimization using computational multi-objective methods., Curr. Opin. Drug Discov. Devel. 10 (2007) 316–324.

[244] K.C.L. Torres, B.R. Souza, D.M. Miranda, a M. Sampaio, R. Nicolato, F.S. Neves, et al., Expression of neuronal calcium sensor-1 (NCS-1) is decreased in leukocytes of schizophrenia and bipolar disorder patients., Prog. Neuropsychopharmacol. Biol. Psychiatry. 33 (2009) 229–34.

[245] A. Ganesan, C.H. Heathcock, - 4 0 2, (1993) 6155–6157.

[246] D.M. Volochnyuk, S. V Ryabukhin, A.S. Plaskon, Y. V Dmytriv, O.O. Grygorenko, P.K. Mykhailiuk, et al., Approach to the library of fused pyridine-4-carboxylic acids by Combes-type reaction of acyl pyruvates and electron-rich amino heterocycles., J. Comb. Chem. 12 (2010) 510–7.

[247] J.C. Sloop, Quinoline formation via a modified Combes reaction: examination of kinetics, substituent effects, and mechanistic pathways, J. Phys. Org. Chem. 22 (2009) 110–117.

[248] T. Mukaiyama, New Synthetic Reactions Based on the Onium Salts of Aza-Arenes [New synthetic methods (29)], Angew. Chemie Int. Ed. English. 18 (1979) 707–721.

[249] C.A.G.N. Montalbetti, V. Falque, Amide bond formation and peptide coupling, Tetrahedron. 61 (2005) 10827–10852.

[250] E. Valeur, M. Bradley, Amide bond formation: beyond the myth of coupling reagents, Chem. Soc. Rev. 38 (2009) 606–631.

[251] M. Wijtmans, D. a Pratt, J. Brinkhorst, R. Serwa, L. Valgimigli, G.F. Pedulli, et al., Synthesis and reactivity of some 6-substituted-2,4-dimethyl-3-pyridinols, a novel class of chain-breaking antioxidants., J. Org. Chem. 69 (2004) 9215–23.

[252] C. Deshayes, M. Chabannet, S. Gelin, Synthesis of some 1-substituted-3-or-5-styrylpyrazoles, J. Heterocycl. Chem. 18 (1981) 1057–1059.

[253] C.I. Chiriac, The direct carboxylation of pyrazoles, Synthesis (Stuttg). (1986) 753–755.

[254] C.D. Edlin, G. Morgans, S. Winks, S. Du, V.M. Avery, S. Wittlin, et al., Identi fi cation and In-Vitro ADME Assessment of a Series of Novel Anti-Malarial Agents Suitable for Hit-to-Lead Chemistry, (2012) 570–573.

[255] S. Grzesiek, A. Bax, Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR, J. Am. Chem. Soc. 114 (1992) 6291–6293.

[256] S. Grzesiek, A. Bax, An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins, J. Magn. Reson. 99 (1992) 201–207.

[257] L.E. Kay, M. Ikura, R. Tschudin, A. Bax, Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins, J. Magn. Reson. 89 (1990) 496–514.

[258] D.R. Muhandiram, L.E. Kay, Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity, J. Magn. Reson. Ser. B. 103 (1994) 203–216.

[259] S. Grzesiek, A. Bax, Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein, J. Magn. Reson. 96 (1992) 432–440.

[260] R.T. Clubb, V. Thanabal, G. Wagner, A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue 1HN, 15N, and 13C′ chemical shifts in 15N□13C-labelled proteins, J. Magn. Reson. 97 (1992) 213–217.

[261] A. Bax, M. Ikura, An efficient 3D NMR technique for correlating the proton and15N backbone amide resonances with the α-carbon of the preceding residue in uniformly15N/13C enriched proteins, J. Biomol. NMR. 1 (1991) 99–104.

[262] B.W. Mcferran, J.L. Weiss, R.D. Burgoyne, Neuronal Ca 2 ⁺ Sensor 1, 274 (1999) 30258–30265.

[263] R.D. Burgoyne, D.W. O'Callaghan, B. Hasdemir, L.P. Haynes, A. V. Tepikin, Neuronal Ca2+-sensor proteins: Multitalented regulators of neuronal function, Trends Neurosci. 27 (2004) 203–209.

[264] M.F. Leopold, J. Urbauer, A.J. Wand, Resonance assignment strategies for the analysis of nmr spectra of proteins, Mol. Biotechnol. 2 (1994) 61–93.

[265] D.S. Wishart, B.D. Sykes, F.M. Richards, The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, Biochemistry. 31 (1992) 1647–1651.

[266] Y.. D.J.. P.V.. M.P.. P.O. Bourne, Immunocytochemical localization and crystal structure of human frequenin (neuronal calcium sensor 1)., J.Biol.Chem. 276 (2001) 11949–11955.

[267] M.. K.V.. P.S.. B.R.. D.J.P.. L.L.Y. Saleem, Neuronal Calcium Sensor (Ncs-1)from Rattus Norvegicus in Complex with D2 Dopamine Receptor Peptide from Homo Sapiens, To Be Publ. (n.d.).

[268] F.A.A. Mulder, D. Schipper, R. Bott, R. Boelens, Altered flexibility in the substrate-binding site of related native and engineered high-alkaline Bacillus subtilisins1, J. Mol. Biol. 292 (1999) 111–123.

[269] S. Campagne, V. Gervais, A. Milon, Nuclear magnetic resonance analysis of protein–DNA interactions, J. R. Soc. Interface. 8 (2011) 1065–1078.

[270] W.B. Turnbull, Divided We Fall? Studying low affinity fragments of ligands by ITC, Microcal Appl. Notes. (2005) 1–10.

[271] J. Lakowicz, ed., Protein Fluorescence, in: Princ. Fluoresc. Spectrosc. SE - 16, Springer US, 2006: pp. 529–575.

[272] J.T. Vivian, P.R. Callis, Mechanisms of tryptophan fluorescence shifts in proteins., Biophys. J. 80 (2001) 2093–2109.

[273] J.G. Topliss, Utilization of operational schemes for analog synthesis in drug design., J. Med. Chem. 15 (1972) 1006–1011.

[274] P. Jeffrey, The Practice of Medicinal Chemistry, Br. J. Clin. Pharmacol. 57 (2004) 662.

[275] T.J. Ritchie, S.J.F. Macdonald, S. Peace, S.D. Pickett, C.N. Luscombe, The developability of heteroaromatic and heteroaliphatic rings – do some have a better pedigree as potential drug molecules than others?, Medchemcomm. 3 (2012) 1062.

[276] R.K. Arvela, N.E. Leadbeater, H.M. Torenius, H. Tye, Rapid cyanation of aryl iodides in water using microwave promotion, Org. Biomol. Chem. 1 (2003) 1119–1121.

[277] G.P. Ellis, T.M. Romney-Alexander, Cyanation of aromatic halides, Chem. Rev. 87 (1987) 779–794.

[278] K. Takagi, T. Okamoto, Y. Sakakibara, S. Oka, PALLADIUM(II) CATALYZED SYNTHESIS OF ARYL CYANIDES FROM ARYL HALIDES, Chem. Lett. 2 (1973) 471–474.

[279] P.G.M. Wuts, T.W. Greene, Greene's protective groups in organic synthesis, John Wiley & Sons, 2006.

[280] K. Takagi, T. Okamoto, Y. Sakakibara, A. Ohno, S. Oka, N. Hayama, Nucleophilic displacement catalyzed by transition metal, III: Kinetic investigation of the cyanation of iodobenzene catalyzed by palladium (II), Bull. Chem. Soc. Jpn. 49 (1976) 3177.

[281] P. Anbarasan, T. Schareina, M. Beller, Recent developments and perspectives in palladium-catalyzed cyanation of aryl halides: synthesis of benzonitriles., Chem. Soc. Rev. 40 (2011) 5049–67.

[282] J. Chatt, B.L. Shaw, 980. Hydrido-complexes of platinum (II), J. Chem. Soc. (1962) 5075–5084.

[283] K.D. Dobbs, W.J. Marshall, V. V Grushin, Why Excess Cyanide Can Be Detrimental to Pd-Catalyzed Cyanation of Haloarenes. Facile Formation and Characterization of [Pd(CN)3(H)]2- and [Pd(CN)3(Ph)]2-, J. Am. Chem. Soc. 129 (2006) 30–31.

[284] S. Erhardt, V. V Grushin, A.H. Kilpatrick, S.A. Macgregor, W.J. Marshall, D.C. Roe, Mechanisms of Catalyst Poisoning in Palladium-Catalyzed Cyanation of Haloarenes. Remarkably Facile C−N Bond Activation in the [(Ph3P)4Pd]/[Bu4N]+ CN- System, J. Am. Chem. Soc. 130 (2008) 4828–4845.

[285] P.E. Maligres, M.S. Waters, F. Fleitz, D. Askin, A highly catalytic robust palladium catalyzed cyanation of aryl bromides, Tetrahedron Lett. 40 (1999) 8193–8195.

[286] S.A. Thorat, D.W. Kang, H. Ryu, M.S. Kim, H.S. Kim, J. Ann, et al., 2-(3-Fluoro-4-methylsulfonylaminophenyl)propanamides as potent TRPV1 antagonists: Structure activity relationships of the 2-oxy pyridine C-region, Eur. J. Med. Chem. 64 (2013) 589–602.

[287] Y. Miyamoto, Y. Banno, T. Yamashita, T. Fujimoto, S. Oi, Y. Moritoh, et al., Discovery of a 3-pyridylacetic acid derivative (TAK-100) as a potent, selective and orally active dipeptidyl peptidase IV (DPP-4) inhibitor, J. Med. Chem. 54 (2011) 831–850.

[288] L.A. Carpino, 1-Hydroxy-7-azabenzotriazole. An efficient peptide coupling additive, J. Am. Chem. Soc. 115 (1993) 4397–4398.

[289] L.A. Carpino, H. Imazumi, A. El-Faham, F.J. Ferrer, C. Zhang, Y. Lee, et al., The Uronium/Guanidinium Peptide Coupling Reagents: Finally the True Uronium Salts, Angew. Chemie Int. Ed. 41 (2002) 441–445.

[290] Solon Mardapittas, Synthesis of novel analogues of a small drug molecule inhibitor ofthe NCS-1/D2 interaction, Liverpool, 2014.

[291] B. Meyer, T. Peters, NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors, Angew. Chemie - Int. Ed. 42 (2003) 864–890.

[292] G. Otting, E. Liepinsh, B. Farmer II, K. Wüthrich, Protein hydration studied with homonuclear 3D1H NMR experiments, J. Biomol. NMR. 1 (1991) 209–215.

[293] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, A. Vulpetti, M. Sundström, Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water*, J. Biomol. NMR. 18 (2000) 65–68.

[294] H.M. Rawel, S.K. Frey, K. Meidtner, J. Kroll, F.J. Schweigert, Determining the binding affinities of phenolic compounds to proteins by quenching of the intrinsic tryptophan fluorescence, Mol. Nutr. Food Res. 50 (2006) 705–713.

[295] D.E. Scott, A.G. Coyne, S. a. Hudson, C. Abell, Fragment-based approaches in drug discovery and chemical biology, Biochemistry. 51 (2012) 4990–5003.

[296] C.A. Lepre, Library design for NMR-based screening, 6 (2001).

[297] N. Okimoto, N. Futatsugi, H. Fuji, A. Suenaga, High-Performance Drug Discovery : Computational Screening by Combining Docking and Molecular Dynamics Simulations, 5 (2009).

[298] F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer, M.D. Brice, J.R. Rodgers, et al., The Protein Data Bank. A computer-based archival file for macromolecular structures., Eur. J. Biochem. 80 (1977) 319–324.

[299] P.A.V. Der Merwe, Surface plasmon resonance in Protein-Ligand Interactions: hydrodynamics and calorimetry., Protein-Ligand Interact. …. (2001) 137–170.

[300] J. Concepcion, K. Witte, C. Wartchow, S. Choo, D. Yao, H. Persson, et al., Label-free detection of biomolecular interactions using BioLayer interferometry for kinetic characterization., Comb. Chem. High Throughput Screen. 12 (2009) 791–800.

[301] S. a I. Seidel, P.M. Dijkman, W. a. Lea, G. van den Bogaart, M. Jerabek-Willemsen, A. Lazic, et al., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions, Methods. 59 (2013) 301–315.

[302] A.K. Kenworthy, Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy, Methods. 24 (2001) 289–296.

[303] D. a. R. Sanders, Protein X-ray Crystallography, Handb. Neurochem. Mol. Neurobiol. (2007) 457 – 465.

[304] S.B. Shuker, P.J. Hajduk, R.P. Meadows, S.W. Fesik, Discovering high-affinity ligands for proteins: SAR by NMR, Science (80-.). 274 (1996) 1531–1534.

[305] A. Bhunia, S. Bhattacharjya, S. Chatterjee, Applications of saturation transfer difference NMR in biological systems, Drug Discov. Today. 17 (2012) 505–513.

[306] C. Dalvit, P. Pevarello, M. Tato, M. Veronesi, a. Vulpetti, M. Sundstrom, Identification of compounds with binding affinity to proteins via magnetization transfer from bulk water, J. Biomol. NMR. 18 (2000) 65–68.

[307] R.M. Keller, K. Wüthrich, Assignment of the heme c resonances in the 360 MHz 1 H NMR spectra of cytochrome c, Biochim. Biophys. Acta (BBA)-Protein Struct. 533 (1978) 195–208.

[308] M. Mayer, B. Meyer, Characterization of ligand binding by saturation transfer difference NMR spectroscopy, Angew. Chemie - Int. Ed. 38 (1999) 1784–1788.

[309] I. Bertini, C. Dalvit, J.G. Huber, C. Luchinat, M. Piccioli, ePHOGSY experiments on a paramagnetic protein: location of the catalytic water molecule in the heme crevice of the oxidized form of horse heart cytochrome c, FEBS Lett. 415 (1997) 45–48.

[310] A. Turnbull, S. Boyd, B. Walse, Fragment-based drug discovery and protein–protein interactions, Res. Reports Biochem. Volume 4 (2014) 13.

[311] L. Zhao, D. Cao, T. Chen, Y. Wang, Z. Miao, Y. Xu, Supporting Information Fragment - based Drug Discovery of 2 - thiazolidinones as Inhibitors of the Histone Reader BRD4 Bromodomain, (2013) 1–8.

[312] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, et al., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, J. Med. Chem. 47 (2004) 1739–1749.

[313] L. Pellegrini, S.Y. David, T. Lo, S. Anand, M. Lee, T.L. Blundell, et al., Insights into DNA recombination from the structure of a RAD51–BRCA2 complex, Nature. 420 (2002) 287–293.

[314] D.E. Scott, M.T. Ehebauer, T. Pukala, M. Marsh, T.L. Blundell, A.R. Venkitaraman, et al., Using a Fragment-Based Approach To Target Protein-Protein Interactions, ChemBioChem. 14 (2013) 332–342.

[315] M. Hohwy, L. Spadola, B. Lundquist, P. Hawtin, J. Dahmén, I. Groth-Clausen, et al., Novel prostaglandin D synthase inhibitors generated by fragment-based drug design, J. Med. Chem. 51 (2008) 2178–2186.

[316] D.M. Bayada, H. Hamersma, V.J. van Geerestein, Molecular diversity and representativity in chemical databases, J. Chem. Inf. Comput. Sci. 39 (1999) 1–10.

[317] D. Gorse, A. Rees, M. Kaczorek, R. Lahana, Molecular diversity and its analysis, Drug Discov. Today. 4 (1999) 257–264.

[318] H.E. Gottlieb, V. Kotlyar, A. Nudelman, NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities In the course of the routine use of NMR as an aid for organic chemistry , a day-to-day problem is the identification of signals deriving from common contaminants literature, J. Org. Chem. 62 (1997) 7512–7515.

[319] A.M. Petros, J. Dinges, D.J. Augeri, S. a Baumeister, D. a Betebenner, M.G. Bures, et al., Discovery of a potent inhibitor of the antiapoptotic protein Bcl-xL from NMR and parallel synthesis., J. Med. Chem. 49 (2006) 656–63.

[320] J.H. Lee, Q. Zhang, S. Jo, S.C. Chai, M. Oh, W. Im, et al., Novel Pyrrolopyrimidine-Based R -Helix Mimetics : Cell-Permeable Inhibitors of Protein - Protein Interactions, (2011) 676–679.

[321] M.E. Martone, V.M. Edelmann, M.H. Ellisman, P. Nef, Cellular and subcellular distribution of the calcium-binding protein NCS-1 in the central nervous system of the rat, Cell Tissue Res. 295 (1999) 395–407.

[322] N.L. Allinger, Y.H. Yuh, J.H. Lii, Molecular mechanics. The MM3 force field for hydrocarbons. 1, J. Am. Chem. Soc. 111 (1989) 8551–8566.

[323] D. Forum, D. Discovery, update discussion forum PET and knockout mice in A “ Rule of Three ” for fragment-based lead, 8 (2003) 876–877.

[324] S.A. Chavez, A.J. Martinko, C. Lau, M.N. Pham, K. Cheng, D.E. Bevan, et al., Development of β -Amino Alcohol Derivatives That Inhibit Toll-like Receptor 4 Mediated Inflammatory Response as Potential Antiseptics, (2011) 4659–4669.

[325] J.D. Sculley, C.S. Hamilton, Some Amide Derivatives of Certain Aminomethylpyridines, J. Am. Chem. Soc. 75 (1953) 3400–3403.

[326] C.D. Edlin, G. Morgans, S. Winks, S. Duffy, V.M. Avery, D. Waterson, et al., Identification and In-Vitro ADME Assessment of a Series of Novel Anti-Malarial Agents Suitable for Hit-to-Lead Chemistry, 3 (1826) 1–34.

[327] P.A. Bhatia, J.F. Daanen, A.A. Hakeem, T. Kolasa, M.A. Matulenko, K.H. Mortell, et al., Acetamides and benzamides that are useful in treating sexual dysfunction, (2004).

[328] O.M. Ohta Shunsaku*, Shimabayashi Akihiro, Aono Mari, A General Convenient One-Pot Procedure for the Conversion of Carboxylic Acids into their t-Butyl Esters which is also Applicable to Aliphatic Carboxylic Acids, Synthesis (Stuttg). (1982) 833–834.

[329] T. Tanaka, H. Sugawara, H. Maruoka, S. Imajo, T. Muto, Discovery of novel series of 6-benzyl substituted 4-aminocarbonyl-1, 4-diazepane-2, 5-diones as human chymase inhibitors using structure-based drug design, Bioorg. Med. Chem. 21 (2013) 4233–4249.

[330] S. Yao, D. Gallenkamp, K. Wölfel, B. Lüke, M. Schindler, J. Scherkenbeck, Synthesis and SERCA activities of structurally simplified cyclopiazonic acid analogues, Bioorg. Med. Chem. 19 (2011) 4669–4678.

[331] D.J. Madar, H. Kopecka, D. Pireh, H. Yong, Z. Pei, X. Li, et al., Discovery of 2-[4-{{2-(2 S, 5 R)-2-Cyano-5-ethynyl-1-pyrrolidinyl]-2-oxoethyl] amino]-4-methyl-1-piperidinyl]-4-pyridinecarboxylic Acid (ABT-279): A Very Potent, Selective, Effective, and Well-Tolerated Inhibitor of Dipeptidyl Peptidase-IV, Useful for the, J. Med. Chem. 49 (2006) 6416–6420.

[332] L.A. Carpino, A. El-Faham, F. Albericio, Racemization studies during solid-phase peptide synthesis using azabenzotriazole-based coupling reagents, Tetrahedron Lett. 35 (1994) 2279–2282.

[333] M. Imazawa, F. Eckstein, Facile synthesis of 2’-amino-2'-deoxyribofuranosylpurines, J. Org. Chem. 44 (1979) 2039–2041.

[334] G.F. Holland, Antidiabetic pyrrolecarboxylic acids, (1981).

[335] S. Caddick, D.B. Judd, A.K. d. K. Lewis, M.T. Reich, M.R.. Williams, A generic approach for the catalytic reduction of nitriles, Tetrahedron. 59 (2003) 5417–5423.

[336] R.D. Burgoyne, D.W. O’Callaghan, B. Hasdemir, L.P. Haynes, A. V Tepikin, Neuronal Ca 2+-sensor proteins: multitalented regulators of neuronal function, Trends Neurosci. 27 (2004) 203–209.

[337] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4 Nature 227: 680–685, Find This Artic. Online. (1970).

[338] T. Yamazaki, W. Lee, C.H. Arrowsmith, D.R. Muhandiram, L.E. Kay, A suite of triple resonance NMR experiments for the backbone assignment of 15N, 13C, 2H labeled proteins with high sensitivity, J. Am. Chem. Soc. 116 (1994) 11655–11666.

[339] E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, Proteins. 59 (2005) 687696.

Appendix

Appendix

The following figures and tables relate to data discussed within the relevant chapters contained within this thesis.

A.1 NMR Spectroscopy: First Generation Compounds

Table A.1.1- **Aliphatic methyl sidechain transfer of assignment.** A table of the amino acid residues of NCS1 for which the assignment of the corresponding aliphatic methyl groups was possible. This table lists those residues that were unaffected by the corresponding inhibitor in relation to their position within the four EF hand motifs of NCS1. With the DMSO spectra as the control, those residues present in each column for each Inhibitor can be deemed unaffected. However those methyl groups missing possibly indicate either an interaction between the inhibitor and the sidechain at that specific residue, or that there are considerable structural changes involving the residues whose resonances were missing. Complete attenuation of a single methyl group for those residues that contain only one, or both for those residues that contain two groups indicate a possible change in structure where the amino acid may have become buried.

(For the purposes of this table residues are labeled in single letter nomenclature a=alpha, b=beta)

	EF-1	EF-2	EF-3	EF-4
DMSO control	V_a12 V_b12 V_a13 V_b13 L_a16 L_b16 T17 T20 T23 K_a25 K_b25 V_a27 V_b27 I_a35 I_b35 K_b36 L_a43 L_b43 A45 I_a51 I_b51	T62 T66 V_a68 V_b68 V_b73 I_a80 I_a86 I_b86 A88 L_b89 V_a91 V_b91 T92	T96 L_a97 L_b97 L101 A104 L_a107 L_b107 L_a110 L_b110 L_a116 L_b116 T117 L_a122 I_a124 I_b124 V_a125 V_b125 A127 I_a128 I_b128	V_a132 V_b132 T135 V_a136 V_b136 L_a138 T144 V_a149 V_b149 I_a152 I_b152 M155 K_a158 L_a164 L_b164 T165 L_a166 L_b166 A175 I_a179 I_b179 V_a180 A182 L_a189 V_a190 V_b190
Inhibitor 1	V_a12 V_b12 V_a13 V_b13 L_b16 T20 T23 K_a25 K_b25 V_a27 I_a35 I_b35 K_b36 L_a43 L_b43 A45 I_b51	T62 T66 V_b73 I_a80 I_a86 I_b86 A88 V_b91	T96 L_a97 L_b110 L_a116 L_b116 T117 L_a122 V_a125 A127 I_b128	V_b132 T135 V_a136 V_b136 T144 V_a149 V_b149 M155 K_a158 L_a164 T165 L_b166 A175 I_b179 V_a180 L_a189 V_a190 V_b190
Inhibitor 2	V_a12 V_b12 V_a13 V_b13 L_a16 T17 T20 T23 K_a25 K_b25 V_a27 I_a35 I_b35 K_b36 L_a43 L_b43 A45	T62 T66 V_a68 V_b68 I_a80 I_a86 I_b86 A88 V_a91 V_b91	T96 L_a97 L_b97 L101 110L_a L_a110 L_b110 L_a116 L_b116 T117 L_a122 I_a124 A127 I_b128	T135 V_a136 V_b136 L_a138 T144 V_a149 V_b149 I_b152 M155 L_b164 T165 L_b166 A175 V_a180 L_a189
Inhibitor 3	V_a12 V_b12 V_a13 V_b13 L_b16 T17 T20 T23 K_a25 K_b25 V_a27 V_b27 I_a35 I_b35 K_b36 L_a43 L_b43 A45 I_a51	T62 T66 V_a68 V_b68 V_b73 I_a80 I_a86 I_b86 A88 L_b89 V_a91 V_b91 T92	T96 L_a97 L_b97 L101 A104 L_a107 L_b107 L_a110 L_b110 L_a116 L_b116 T117 L_a122 I_a124 I_b124 V_a125 V_b125 A127 I_b128	V_a132 V_b132 T135 V_a136 V_b136 L_a138 T144 V_a149 I_b152 M155 K_a158 L_a164 L_b164 L_a166 L_b166 A175 I_a179 I_b179 V_a180 A182 L_a189 V_a190
Inhibitor 4	V_a12 V_b12 V_a13 V_b13 L_b16 T17 T20 T23 K_a25 K_b25 V_a27 V_b27 I_a35 I_b35 K_b36 L_a43 L_b43 A45 I_a51 I_b51	T62 T66 V_a68 V_b68 V_b73 I_a80 I_a86 I_b86 A88 L_b89 V_a91 V_b91 T92	T96 L_a97 L_b97 L101 A104 L_a107 L_b107 L_a110 L_b110 L_a116 L_b116 T117 L_a122 I_a124 I_b124 V_a125 V_b125 A127 I_a128	V_a132 135 V_a136 V_b136 L_a138 V_a149 V_b149 I_b152 M155 K_a158 L_a164 L_b164 T165 L_a166 L_b166 A175 I_a179 I_b179 V_a180 A182 L_a189 V_a190
Inhibitor 5	V_a12 V_b12 V_a13 V_b13 L_b16 T17 T20 T23 K_a25 K_b25 V_a27 V_b27 I_a35 I_b35 K_b36 L_a43 L_b43 A45 I_a51 I_b51	T62 T66 V_a68 V_b68 V_b73 I_a80 I_a86 I_b86 A88 L_b89 V_a91 V_b91 T92	T96 L_a97 L_b97 L101 A104 L_a107 L_b107 L_a110 L_b110 L_a116 L_b116 T117 L_a122 I_a124 I_b124 V_a125 V_b125 A127 I_a128 I_b128	V_a132 135 V_a136 V_b136 L_a138 T144 V_a149 V_b149 I_b152 M155 L_a164 L_b164 T165 L_a166 L_b166 A175 I_a179 I_b179 V_a180 A182

Table A.1.2- Aromatic methyl sidechain transfer of assignment. A table depicting the transfer of assignment of the aromatic sidechain residues of NCS1 to the DMSO control and subsequent comparison between those residues assigned for the control spectra with those affected by the individual inhibitors. The results indicate a large change in the regions EF-2 and EF-4 of NCS1, an effect caused by all five inhibitors, this is noted as the change in the individual spectra meant that the residues could be assigned. Inhibitor 1 appears to cause the greatest change in the aromatic amino acid residues of NCS1, such that the transfer of assignment for the majority of residues was impossible.

(*It should be noted that the transfer of assignment between the control and the individual inhibitors was not possible if the peaks were attenuated, broadened or if they were untraceable. For the purposes of this table residues are labeled in single letter nomenclature a=alpha, b=beta).*

	EF-1	EF-2	EF-3	EF-4
DMSO control	Y_a21 Y_b21 Y_a31 Y_a52 Y_b52	F_a67 F_b67 F_a82 F_b82 F_a85 F_b85	W103 Y_a108 Y_b108 Y_a115 Y_b115	Y_a129 Y_b129 F153 F169
Inhibitor 1	Y_a21	F_b82	Y_a115	
Inhibitor 2	Y_a21 Y_b21 Y_a31	F_a82	Y_a115 Y_b115	Y_a129
Inhibitor 3	Y_a21 Y_b21 Y_a31 Y_b52		Y_b108 Y_a115 Y_b115	
Inhibitor 4	Y_a21 Y_b21 Y_b52		Y_a108 Y_b108 Y_b115	Y_a129
Inhibitor 5	Y_a21 Y_b21 Y_a31 Y_a52 Y_b52		Y_a108 Y_b108 Y_a115 Y_b115	Y_b129

STD and waterLOGSY spectra acquired for the first generation inhibitors 1, 2 & 3, see Chapter 2.3 and 3.3 for further details.

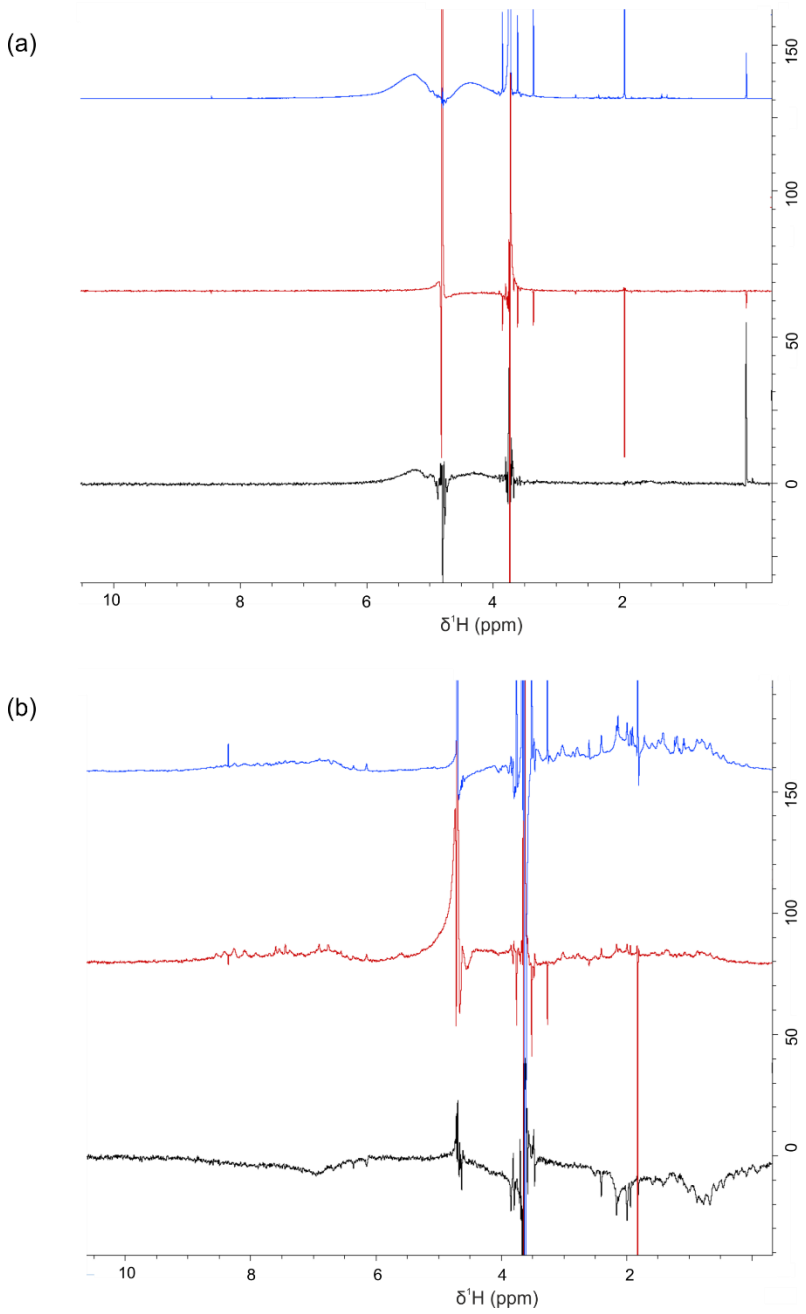


Figure A.1.1- -Inhibitor 1 Ligand screening spectra. **(a)** Buffer standard 500μM Inhibitor 1 in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. **(b)** Binding screen 500 μM Inhibitor 1 & NCS1 50 μM in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. Comparison between the buffer controls (a) and the protein containing experiments (b) are obscured by the reduced solubility of the inhibitor meaning that concentration of ligand in solution is significantly less than required.

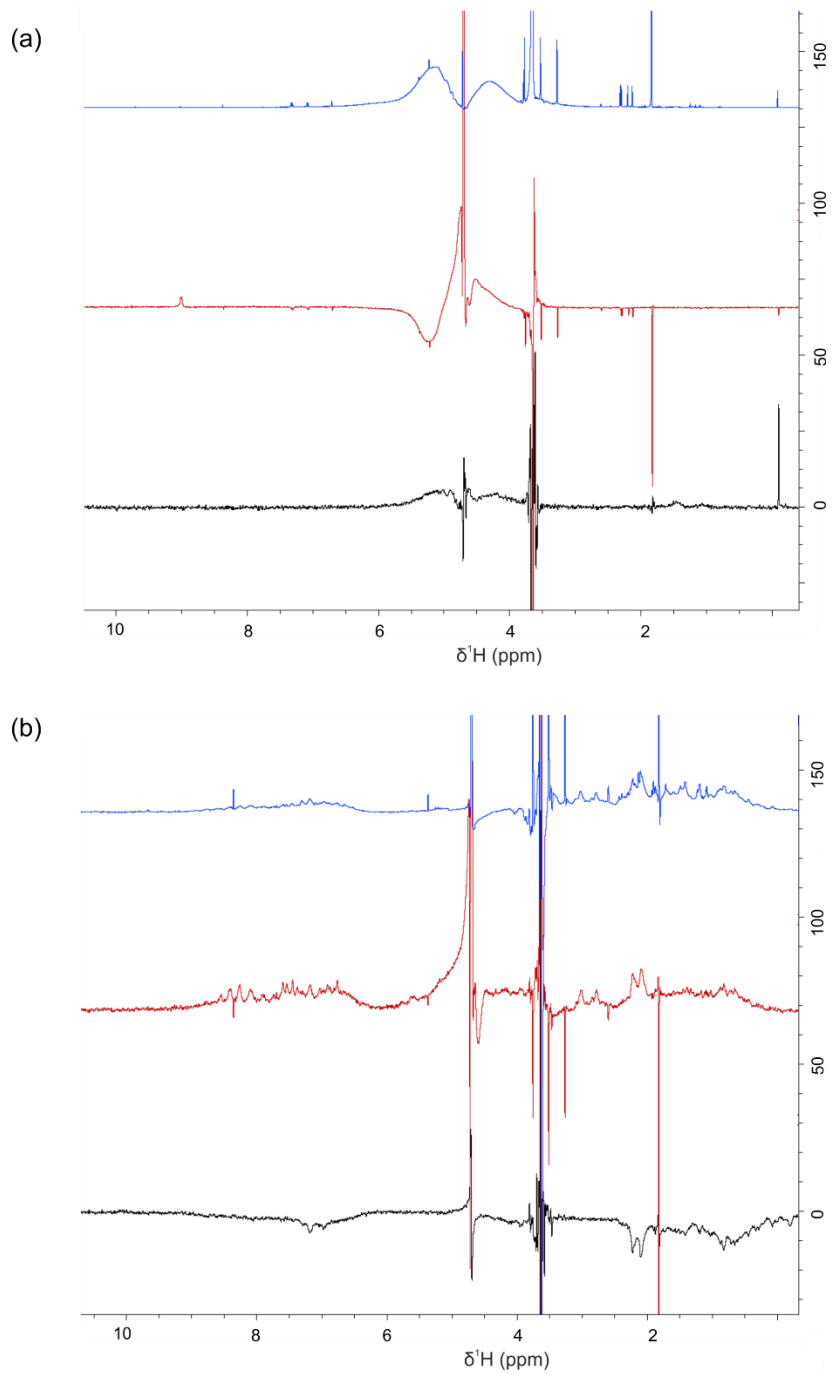
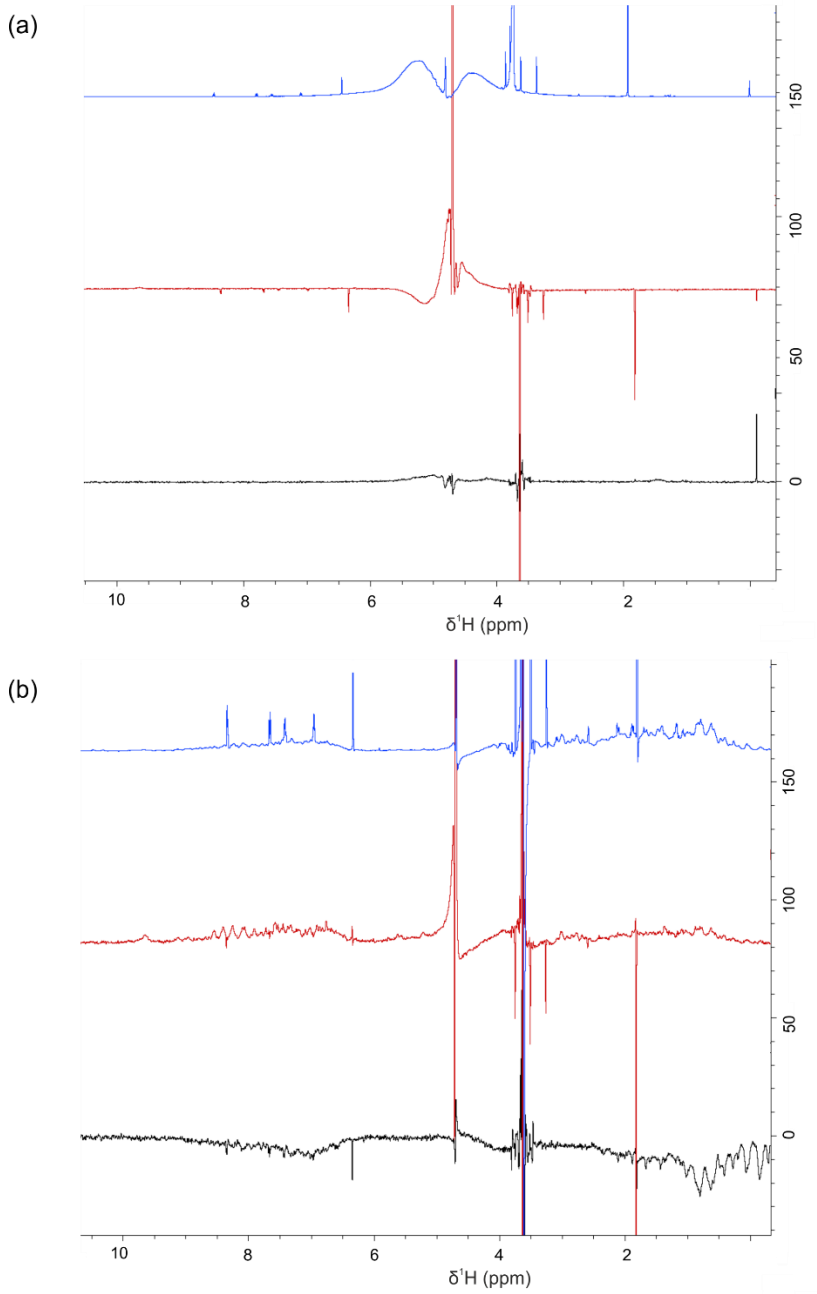


Figure A.1.2- -Inhibitor 2 Ligand screening spectra. **(a)** Buffer standard 500μM Inhibitor 2 in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. **(b)** Binding screen 500 μM Inhibitor 2 & NCS1 50 μM in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. Comparison between the buffer controls (a) and the protein containing experiments (b) are obscured by the reduced solubility of the inhibitor meaning that concentration of ligand in solution is significantly less than required.



Inhibitor 3

Figure A.1.3- -Inhibitor 3 Ligand screening spectra. (a) Buffer standard 500μM Inhibitor 3 in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. **(b)** Binding screen 500 μM Inhibitor 3 & NCS1 50 μM in NMR buffer; 1-D ^{1}H spectra blue, 1-D ^{1}H waterLOGSY spectra red and 1-D ^{1}H STD black. Comparison between the buffer controls (a) and the protein containing experiments (b) are obscured by the reduced solubility of the inhibitor meaning that concentration of ligand in solution is significantly less than required.

A.2 First Generation Isothermal Titration Calorimetry

A.2.1 One Set of Sites fitting functions [212]

The equation used for the fitting of ITC data in Chapter 2.3 and 3.3 is derived from the following equations:

$$K = \frac{\theta}{(1-\theta)[x]} \qquad \text{A.1}$$

$$\mathrm{Xt} = [\mathrm{x}] + n\theta Mt \qquad \text{A.2}$$

$$\theta^2 - \theta\left[1 + \frac{\mathrm{Xt}}{nMt} + \frac{1}{nKMt} + \frac{Xt}{nMt}\right] = 0 \qquad \text{A.3}$$

$$Q = n\theta Mt\Delta HVo \qquad \text{A.4}$$

$$Q = \frac{nMt\Delta HVo}{2}[1 + \frac{\mathrm{Xt}}{nMt} + \frac{1}{nKMt} - \sqrt{(\frac{\mathrm{Xt}}{nMt} + \frac{1}{nKMt})^2 - \frac{4Xt}{nMt}}] \qquad \text{A.5}$$

K = binding constant
n = number of sites
V_o = active cell volume
M_t = bulk concentration of macromolecule in V_o
[m] = free concentration of macromolecule in V_o
X_t = bulk concentration of ligand
[x] = free concentration of ligand
Θ = fraction of sites occupied by the ligand x
Q = total heat content in V_o
ΔQ_i = heat released from the i^{th} injection
ΔH = molar heat of ligand binding

Fitting of the experimental data involves the following processes:

1) Initial estimates by origin of the values for n, K , and ΔH
2) Calculation of ΔQ(i) for each individual injection, these values are subsequently compared with the measured heat measured for the corresponding experimental injection
3) Improvement of n, K , and ΔH by standard Marquardt methods
4) Continuous iteration of the procedure (1, 2 & 3) until there is no improvement in the values.

A.2.2 Inhibitor 1

Data expanded from Chapter 2, Section 2.3.6

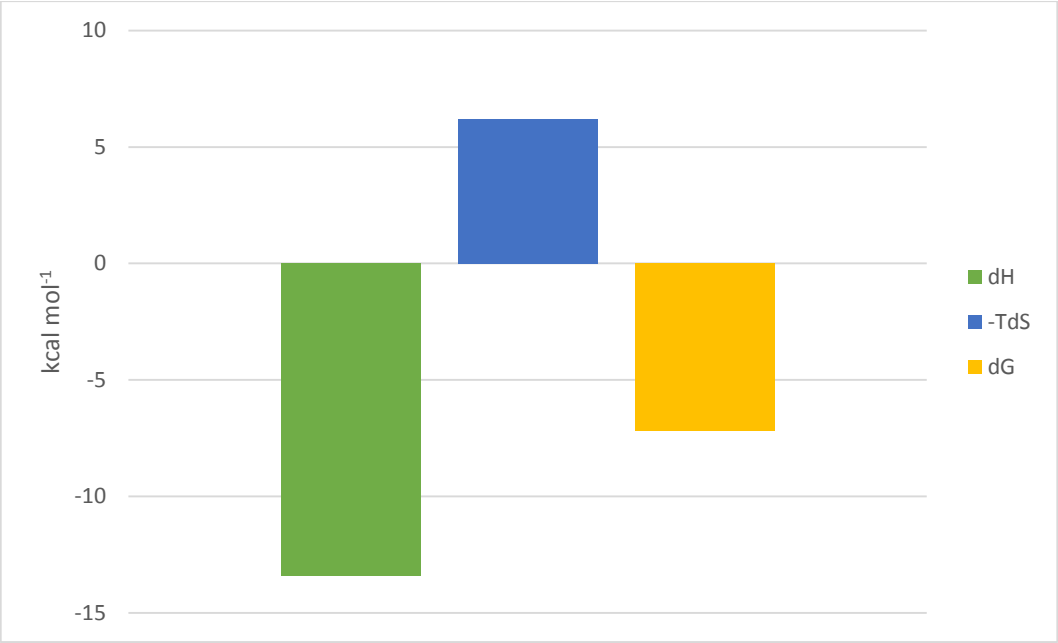


Figure A.2.1- The thermodynamic parameters (dH= ΔH green, -TdS= -TΔS blue and dG= ΔG yellow) obtained from the ITC experiment of the binding interaction between NCS1 400μM and Inhibitor 1 50 μM. When determining ΔG the most influential value is ΔH, ideally it should be large and negative, -TΔS contributes substantially less to the value of ΔG and the binding affinity and tends to be small and positive or small and negative.

Table A.2.1- A table showing the thermodynamic properties from the duplicate ITC experiments of 50 μM Inhibitor 1 with NCS1 400μM. It can be seen that the calculated values do not match those that were obtained from the duplicate experiments and so it can be concluded that the solubility difficulties of Inhibitor 1 prevent any reliable determination of binding affinity.

NCS1 400μM: Inhibitor 1 50μM	**ΔH (kcal mol^{-1})**	**-TΔS (kcal mol-1)**	**ΔG (kcal mol-1)**	**K_D μM**	**Chi^2**
Data adjusted in Origin	-13.4 (±209.5)	6.2	-7.2	5.5	69790
1	-1.8 (±148.2)	2.6	-4.4	3.90E+06	17230
2	-1.9 (±199.4)	-3.8	5.7	60.9	68980

A.2.3 Inhibitor 2

Table A.2.2- A table showing the thermodynamic properties from the ITC experiment of 1 mM NCS1 and 100 µM Inhibitor 2. The value of each property and the calculated error for ΔH are extremely large, indicating a poor goodness of fit. This could be due to the possible precipitation of the inhibitor observed during the experiment and in the related isotherm, hence making the derived K_D unrealistic.

Experiment Title	ΔH (kcal mol^{-1})	-TΔS (kcal mol-1)	ΔG (kcal mol-1)	K_D µM	Chi^2
NCS1 1mM: Inhibitor 2 100µM	-4.708E5 (±1.003E11)	4.71E+08	-4.70E+08	179.5	252000

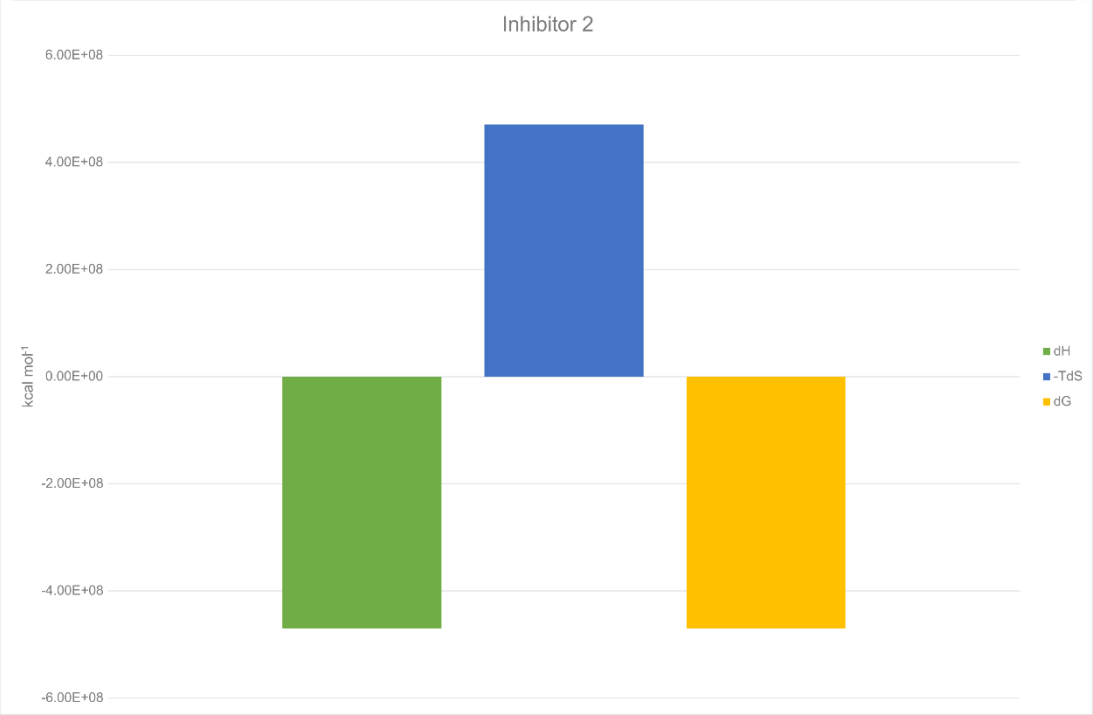


Figure A.2.2- The thermodynamic parameters (dH= ΔH green, -TdS= -TΔS blue and dG= ΔG yellow) obtained from the ITC experiment of the binding interaction between NCS1 1mM and Inhibitor 1 100 µM. When determining ΔG the most influential value is ΔH, ideally it should be large and negative, -TΔS contributes substantially less to the value of ΔG and the binding affinity and tends to be small and positive or small and negative.

A.3 Second generation Computational Design-

The compounds designed as analogues of Inhibitor 2 for the second generation (see Chapter 3 for further details).

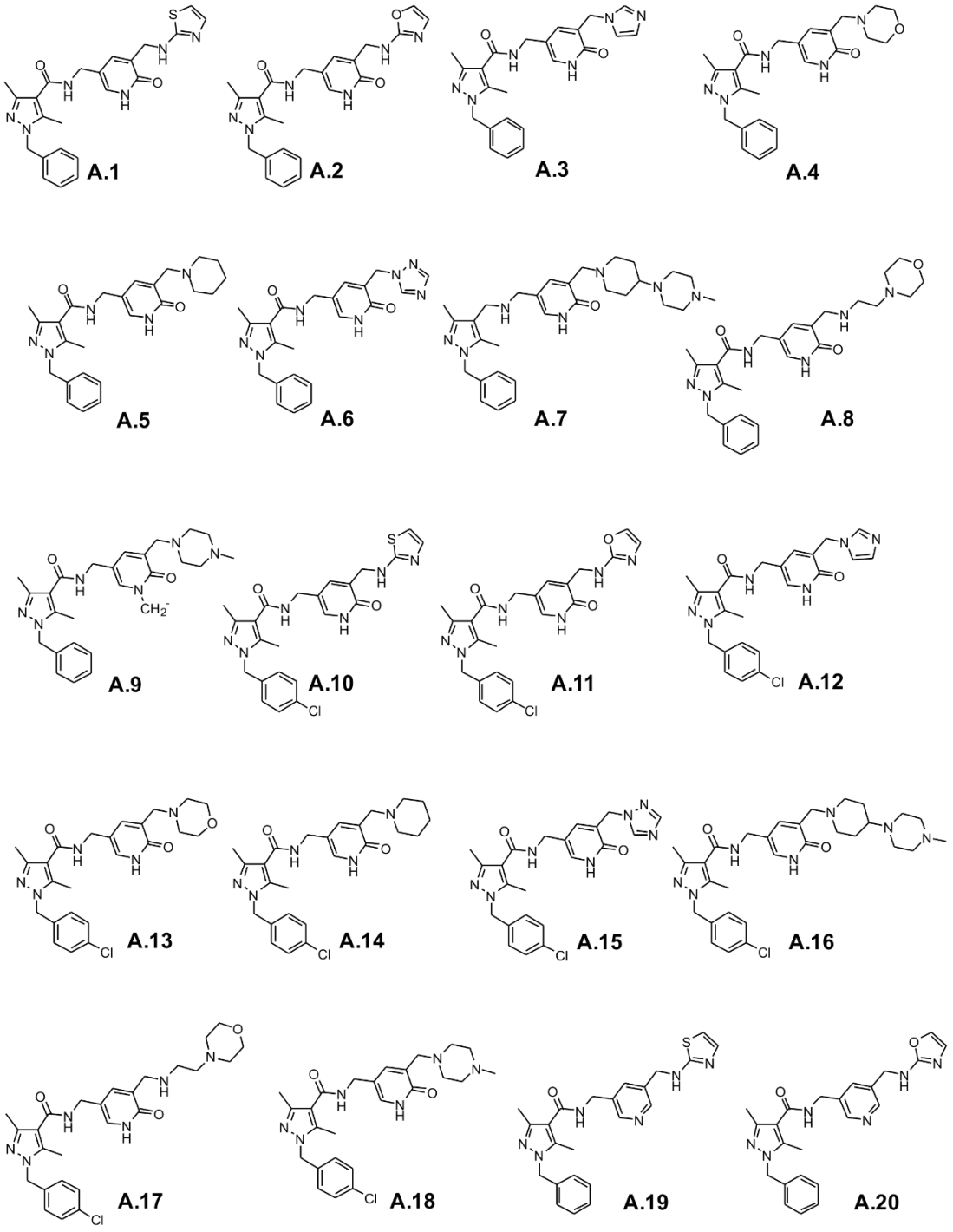


Figure A.3.1- The chemical structures of Inhibitor 2 analogues molecules **A.1** - **A.20** that were subsequently screened using the computational techniques discussed in Chapter 3.1 Second Generation Inhibitor design.

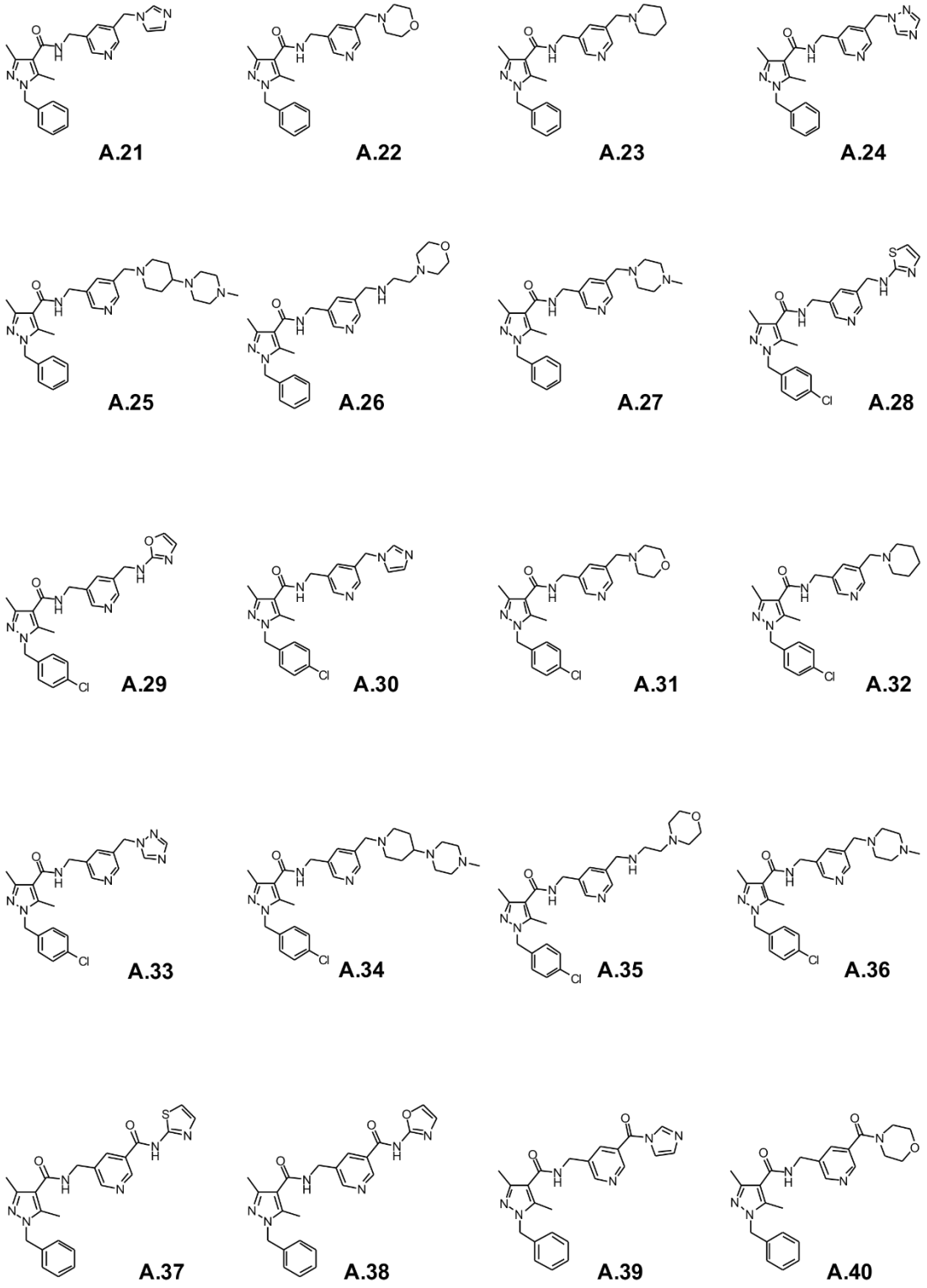


Figure A.3.2- The chemical structures of Inhibitor 2 analogues molecules **A.21** – **A.40** that were subsequently screened using the computational techniques discussed in Chapter 3.1 Second Generation Inhibitor design.

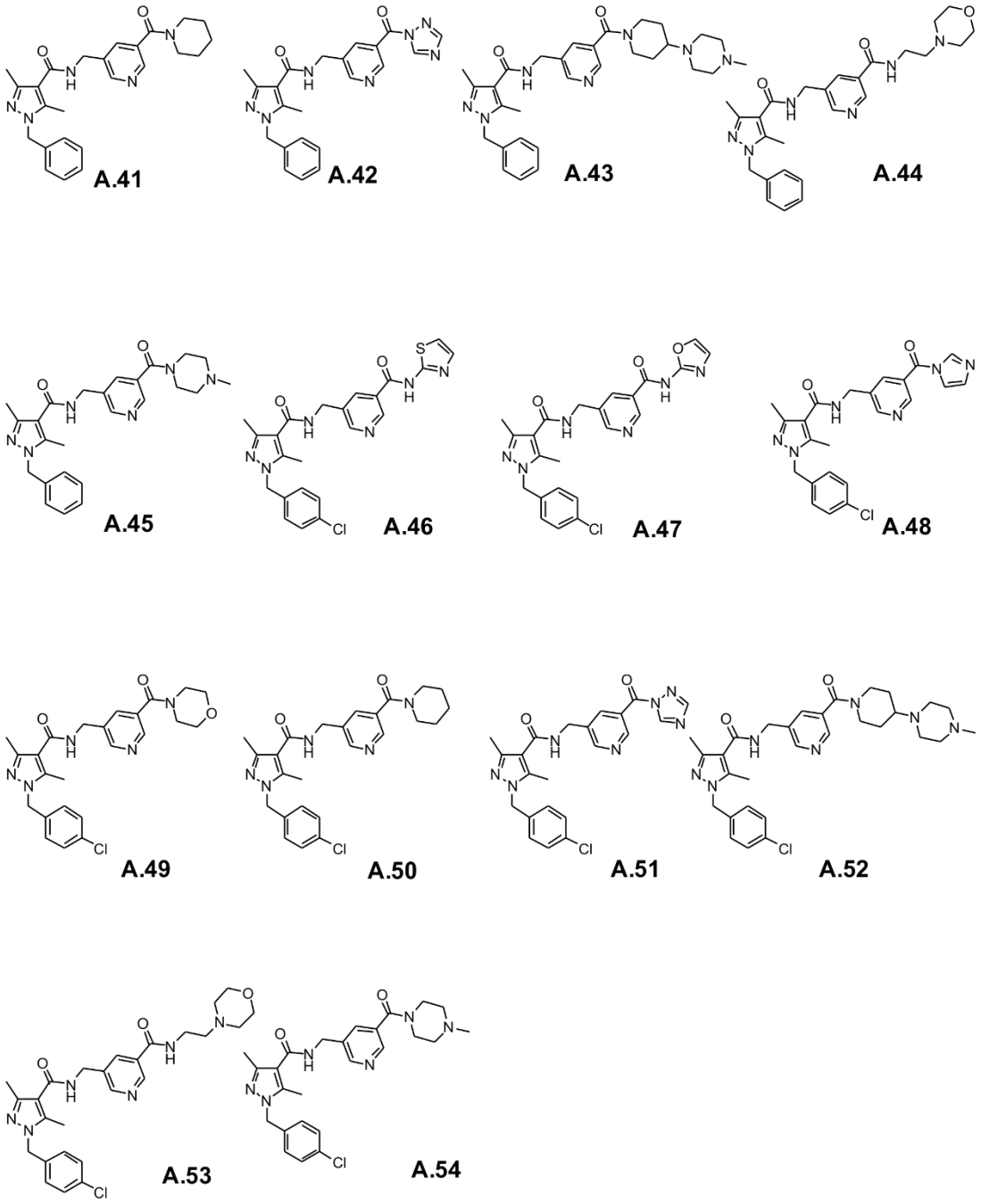


Figure A.3.3- The chemical structures of Inhibitor 2 analogues molecules **A.45** – **A.54** that were subsequently screened using the computational techniques discussed in Chapter 3.1 Second Generation Inhibitor design.

Table A.3.1- Second Generation Inhibitor Properties. The calculated ADMET properties for the Inhibitor 2 analogues molecules **A.1** - **A.54** before any filters were applied to select the most suitable candidates for synthesis. This table includes the Goldscore and ligand efficiency values generated from the docking analysis (Page 1 of 3).

Name	Molecular solubility	Molecular Surface Area	Num H Acceptors	Num H Donors	ALogP	Molecular Weight	Average Goldscore	Ligand Efficiency
Molecule **A.1**	-6.286	431.24	5	3	1.439	448.54	41.99	0.093614839
Molecule **A.2**	-5.744	421.89	5	3	0.88	432.48	41.42	0.095773215
Molecule **A.3**	-5.801	410.48	4	2	0.726	416.48	43.20	0.103726469
Molecule **A.4**	-4.692	430.22	5	2	0.811	435.52	37.94	0.087114254
Molecule **A.5**	-5.732	431.57	4	2	2.04	433.55	39.43	0.090946834
Molecule **A.6**	-5.445	409.37	5	2	0.599	417.46	44.17	0.105806544
Molecule **A.7**	-4.803	533.51	6	2	1.191	531.69	38.97	0.073294589
Molecule **A.8**	-5.178	477.82	6	3	0.413	478.59	36.53	0.076328381
Molecule **A.9**	-4.42	452.88	5	2	1.08	448.56	39.61	0.088304798
Molecule **A.10**	-7.065	454.31	5	3	2.103	482.99	42.88	0.088780306
Molecule **A.11**	-6.524	444.96	5	3	1.545	466.92	40.24	0.086181787
Molecule **A.12**	-6.581	433.55	4	2	1.39	450.92	44.34	0.098332298
Molecule **A.13**	-5.472	453.29	5	2	1.475	469.96	47.04	0.100093625
Molecule **A.14**	-6.511	454.64	4	2	2.704	467.99	42.39	0.090578859
Molecule **A.15**	-6.225	432.44	5	2	1.263	451.91	48.54	0.107410768
Molecule **A.16**	-5.585	556.58	6	2	1.855	566.14	39.70	0.070123998
Molecule **A.17**	-5.956	500.9	6	3	1.078	513.03	44.50	0.086739567
Molecule **A.18**	-5.202	475.95	5	2	1.745	483.01	41.44	0.085795325
Molecule **A.19**	-7.179	422.530	5	2	2.521	432.540	43.96	0.101632219

Table A.3.1- (Page 2 of 3).

Molecule **A.20**	-6.546	413.170	5	2	1.963	416.480	42.97	0.103174222
Molecule **A.21**	-6.612	401.760	4	1	1.809	400.480	42.71	0.106647024
Molecule **A.22**	-5.640	421.500	5	1	1.893	419.520	36.70	0.087480931
Molecule **A.23**	-6.746	422.850	4	1	3.123	417.550	38.43	0.092036882
Molecule **A.24**	-6.197	400.660	5	1	1.681	401.460	43.71	0.108877597
Molecule **A.25**	-5.824	524.790	6	1	2.274	515.690	45.36	0.087959821
Molecule **A.26**	-6.142	469.110	6	2	1.496	462.590	42.58	0.092046953
Molecule **A.27**	-5.372	444.160	5	1	2.163	432.560	39.20	0.090623266
Molecule **A.28**	-7.951	445.600	5	2	3.186	466.990	45.46	0.097346838
Molecule **A.29**	-7.320	436.240	5	2	2.628	450.920	41.32	0.091634880
Molecule **A.30**	-7.386	424.830	4	1	2.473	434.920	46.69	0.107353076
Molecule **A.31**	-6.414	444.570	5	1	2.558	453.960	45.59	0.100427350
Molecule **A.32**	-7.518	445.920	4	1	3.787	451.990	36.92	0.081683223
Molecule **A.33**	-6.972	423.730	5	1	2.346	435.910	44.83	0.102842330
Molecule **A.34**	-6.600	547.860	6	1	2.938	550.140	41.59	0.075598938
Molecule **A.35**	-6.914	492.180	6	2	2.161	497.030	42.82	0.086151741
Molecule **A.36**	-6.148	467.240	5	1	2.827	467.010	37.14	0.079527205
Molecule **A.37**	-6.824	430.670	5	2	1.914	446.520	38.46	0.086132760
Molecule **A.38**	-6.357	421.310	5	2	1.356	430.460	41.24	0.095804488
Molecule **A.39**	-6.395	408.770	5	1	1.674	414.460	41.68	0.100564590
Molecule **A.40**	-5.292	430.240	5	1	1.234	433.500	38.12	0.087935409
Molecule **A.41**	-6.298	431.590	4	1	2.463	431.530	42.23	0.097861099
Molecule **A.42**	-6.093	407.670	6	1	1.546	415.450	38.03	0.091539295
Molecule **A.43**	-5.397	533.530	6	1	1.614	529.680	47.58	0.089827821

Table A.3.1- (Page 3 of 3).

Molecule **A.44**	-5.695	477.340	6	2	1.166	476.570	47.08	0.098789265
Molecule **A.45**	-5.052	452.900	5	1	1.503	446.540	39.18	0.087741300
Molecule **A.46**	-7.598	453.740	5	2	2.578	480.970	41.62	0.086533464
Molecule **A.47**	-7.133	444.390	5	2	2.020	464.900	42.90	0.092277909
Molecule **A.48**	-7.170	431.850	5	1	2.338	448.900	44.44	0.098997550
Molecule **A.49**	-6.066	453.310	5	1	1.898	467.950	37.36	0.079837589
Molecule **A.50**	-7.071	454.660	4	1	3.128	465.980	45.44	0.097514915
Molecule **A.51**	-6.869	430.740	6	1	2.211	449.890	39.13	0.086976817
Molecule **A.52**	-6.174	556.600	6	1	2.279	564.120	50.27	0.089112246
Molecule **A.53**	-6.468	500.410	6	2	1.831	511.020	40.80	0.079840319
Molecule **A.54**	-5.829	475.970	5	1	2.168	480.990	41.02	0.085282438

A.4 Second Generation Isothermal Titration Calorimetry

The isotherm is expanded from Chapter 3.3.2

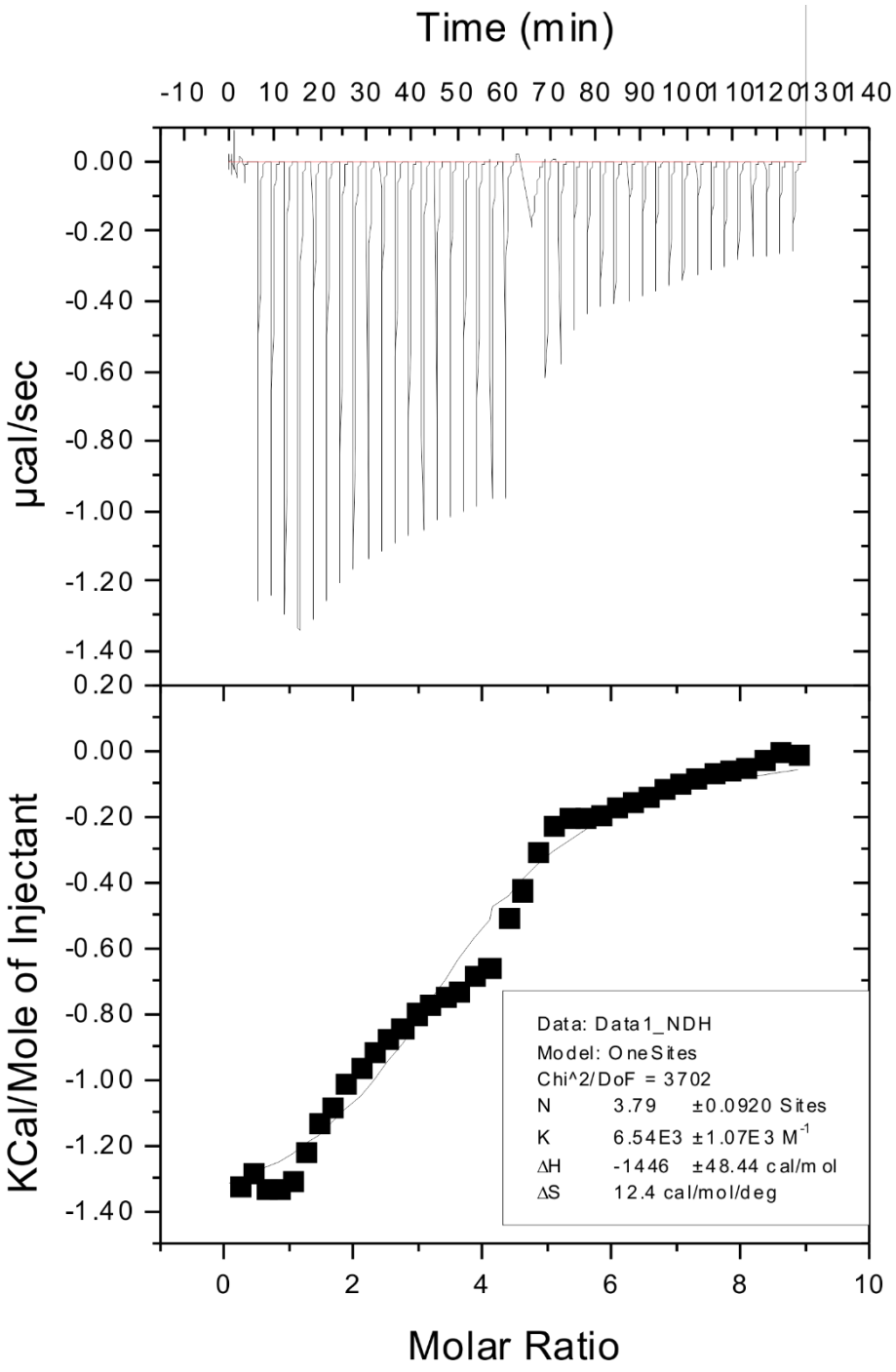
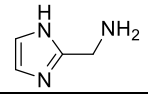
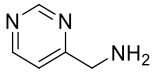
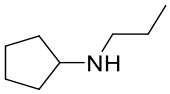
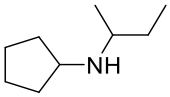
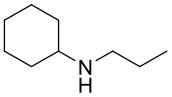
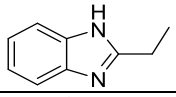
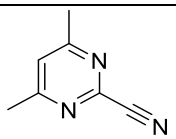
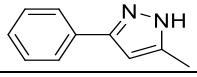
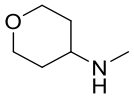
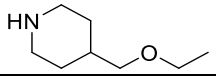
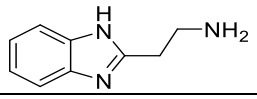
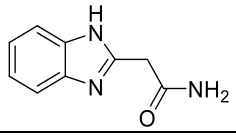
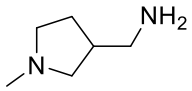
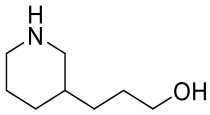
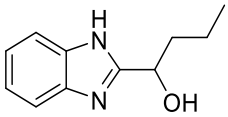
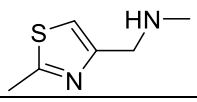
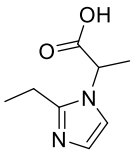
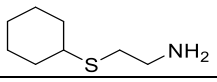
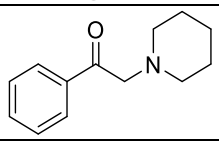
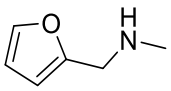
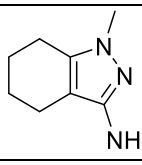
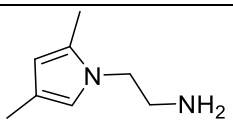
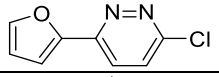
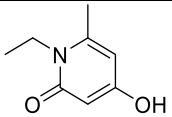
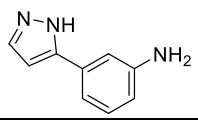


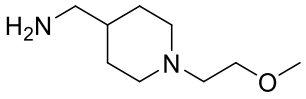
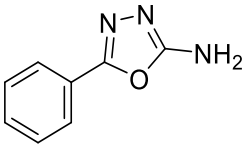
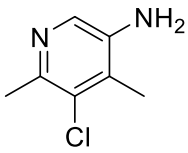
Figure A.4.1- Binding affinity determination- Isothermal Titration Calorimetry Curve - ITC Isotherm (top) and resultant curve (bottom) from the two sequential experiments of Inhibitor 5 8 mM titrated in to NCS1 400 µM at 10°C, 8% DMSO, 50 mM tris HCl, 50 mM NaCl, 5 mM $CaCl_2$ pH 7.5. A large change in heat between the the end of the first experiement and the start of the second, meaning that the goodness of fit (Chi^2) was extremly large and unfavourable. Fitting of the curve was conducted in Origin 7 using the single set of sites curve fitting model produced a binding model with a derived binding affinity K_D of 152.9 µM undesiarable enthalpic values as follows ΔH= -1.446 Kcal, $-T\Delta S$= 2.604Kcal ΔG=-4.05 Kcal.

A.5 Small Molecule Compound Library experimental spectra

Table A.5.1- A table containing the details of the 28 chemical compounds within the chemical library, including the individual chemical structures of each compound, with details of the results of the experiments performed; 1-D ^{1}H saturation transfer difference (STD), 1-D ^{1}H waterLOGSY and 2-D ^{1}H ^{15}N HSQC. If the compound did not appear to have an effect upon the protein in the initial STD screen then the secondary waterLOGSY and HSQC experiments were not applicable (N/A). All data was collected on a 600 MHz spectrometer equipped with a cryoprobe at 298K. The initial screen used the STD experiment, the NMR sample 550µL was composed of NCS1 20µM (v/v), small molecule 1mM (v/v) dissolved in 100% DMSO, with 10% (v/v) $^{2}D_2O/H_2O$, further details of the experiment can be found in 6.3.2.10. The indication of an interaction between the protein and small molecule was observed through the re-occurrence of the ligand signal in the STD experiment. This was then confirmed by the use of a WaterLOGSY experiment, where the corresponding ligand peak intensity would be reduced, disappear completely or become positive (the opposite to the STD spectra). From the initial screen those compounds that appeared to undergo an interaction with NCS1 were then tested further, increasing the concentration of NCS1 in the NMR sample to 50µM the STD and waterLOGSY experiments were repeated and a further ^{1}H ^{15}N HSQC spectra was collected to observe those residues of NCS1 that were affected by the small molecule.

Compound Name		Indication of binding through STD	Indication of binding through WaterLOGSY	Indication of binding through HSQC
4.14		No	No	No
4.15		No	N/A	N/A
4.16		Possibly weak Aliphatic region	Yes weak binding	Yes weak binding
4.17		Possible weak binding	No	No
4.18		No	N/A	N/A
4.19		No	N/A	N/A
4.20		Possibly weak binding	No	No
4.21		Yes	Yes	Yes
4.22		Possible although weak	No	No
4.23		Possibly weak	No	No

4.24		Possibly weak	Yes weak binding	Yes weak binding
4.25		No	N/A	N/A
4.26		No	N/A	N/A
4.27		No	N/A	N/A
4.28		Yes	Yes	Yes
4.29		No	N/A	N/A
4.30		Possibly weak	No	No
4.31		Possibly weak	Yes weak binding	Yes weak binding
4.32		Possibly weak	Yes weak binding	Yes weak binding
4.33		Possibly Weak	No	No
4.34		Possible but weak	No	No
4.35		Yes	No	No
4.36		Yes	Yes	Yes
4.37		Possibly weak	No	No
4.38		Possibly weak aromatic region	No	No

4.39		Possibly weak aromatic region	No	No
4.40		Yes	Yes	Yes
4.41		Yes	Yes	Yes

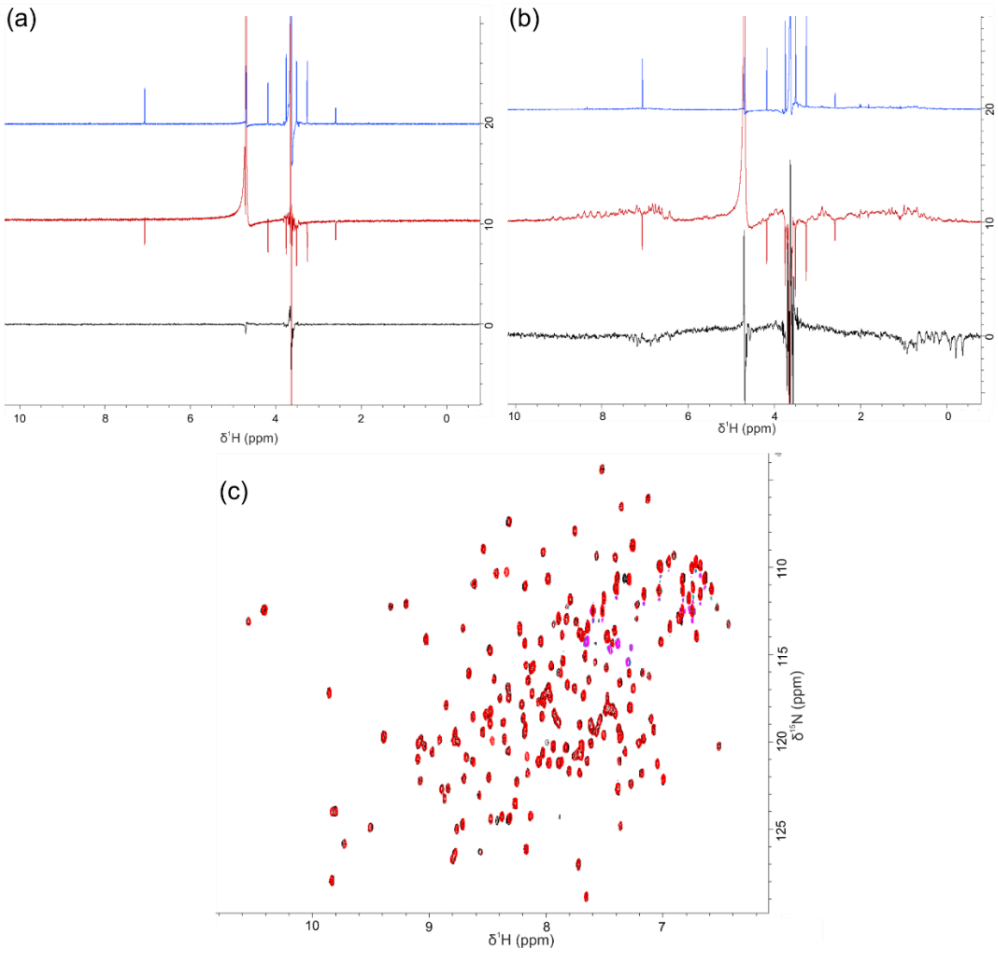


Figure A.5.1- NMR spectra collected for Compound 4.14. (a) The buffer standard, 1 mM **4.14** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.14** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. It was thought a possible indication of a weak interaction was observed in these experiments. However **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50µM and

DMSO 1% (black) with the spectrum of the ligand 1mM in DMSO 1% and NCS1 50 µM red indicated no interaction.

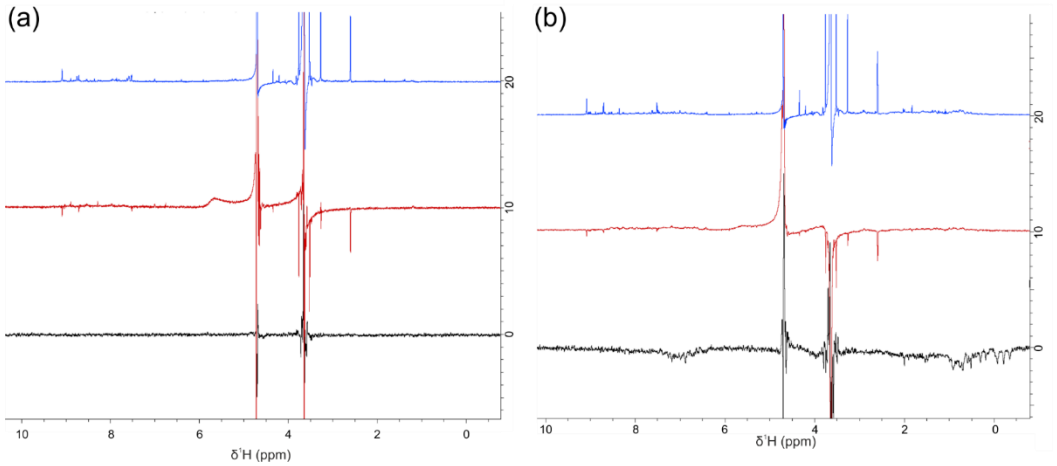


Figure A.5.2- NMR spectra collected for Compound 4.15. (a) The buffer standard, 1 mM **4.15** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.15** and 20 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

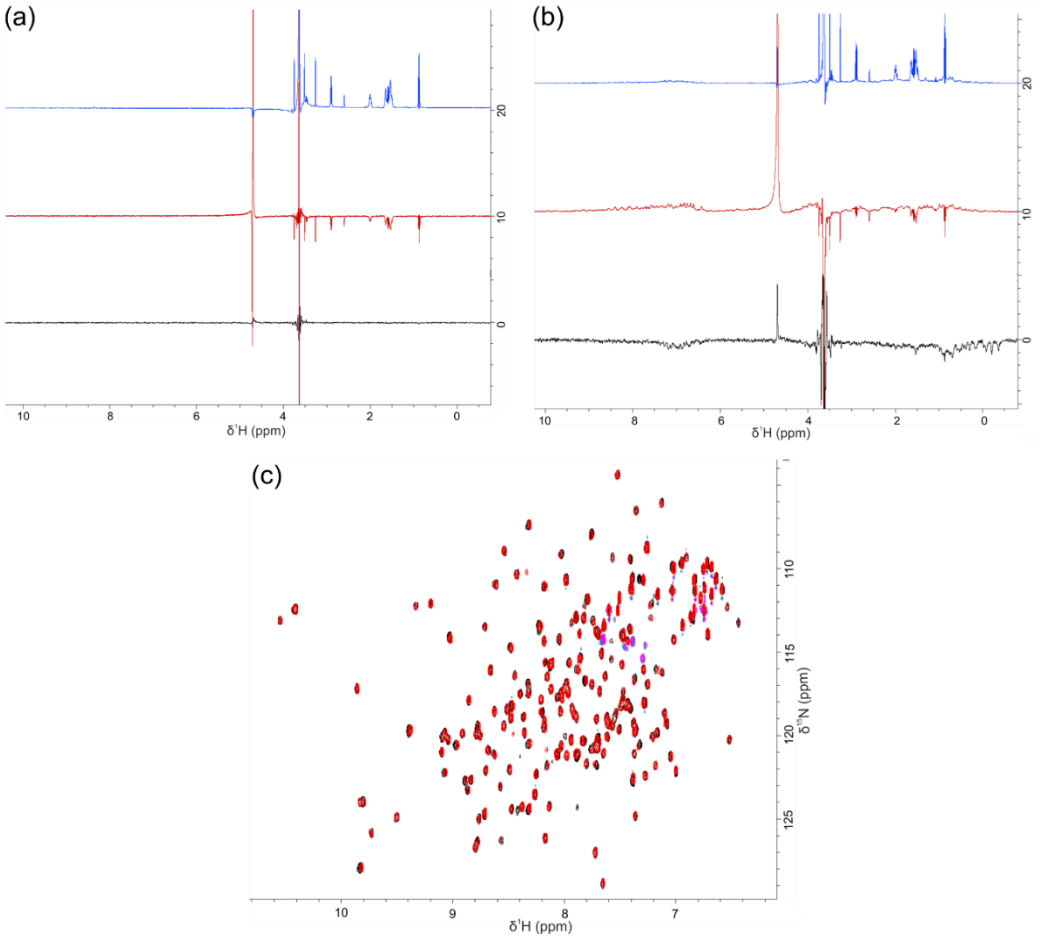


Figure A.5.3- NMR spectra collected for Compound 4.16. (a) The buffer standard, 1 mM **4.16** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.16** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to be observed in the aliphatic region of the

STD spectra. This was confirmed by **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, whereby some of the residues appear partially perturbed

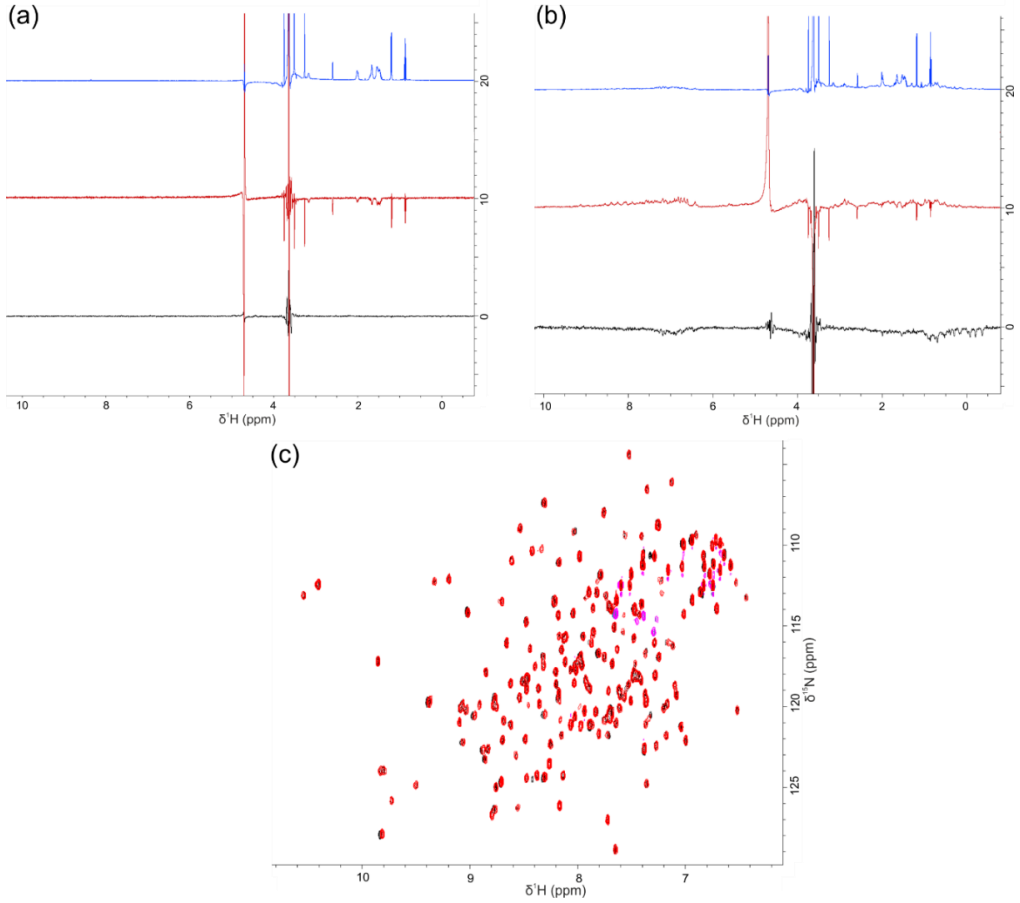


Figure A.5.4- NMR spectra collected for Compound 4.17. **(a)** The buffer standard, 1 mM **4.17** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.17** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.17** and NCS1, in the aliphatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, both suggested that this was a false result.

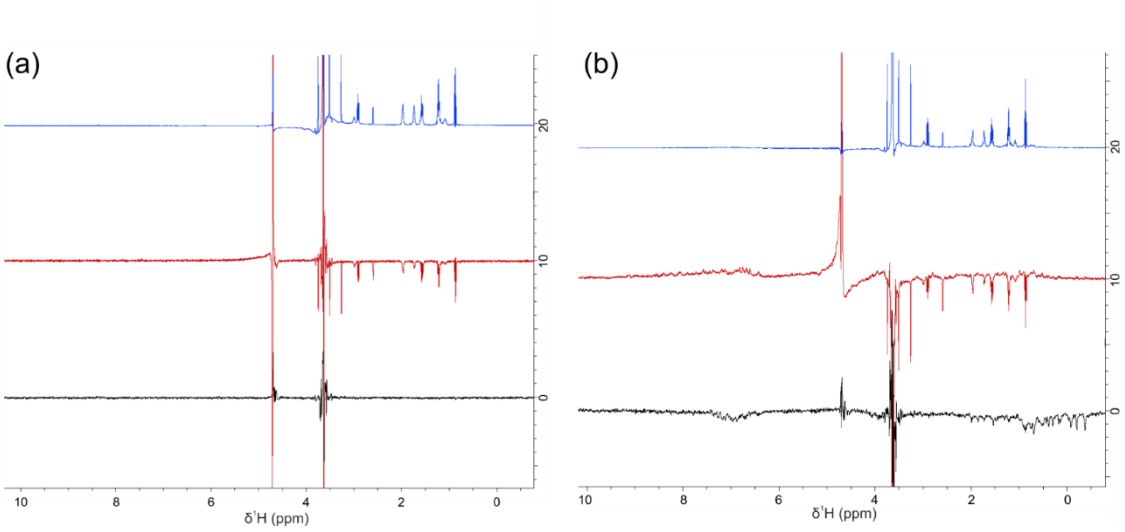


Figure A.5.5- NMR spectra collected for Compound 4.18. **(a)** The buffer standard, 1 mM **4.18** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.18** and 20 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

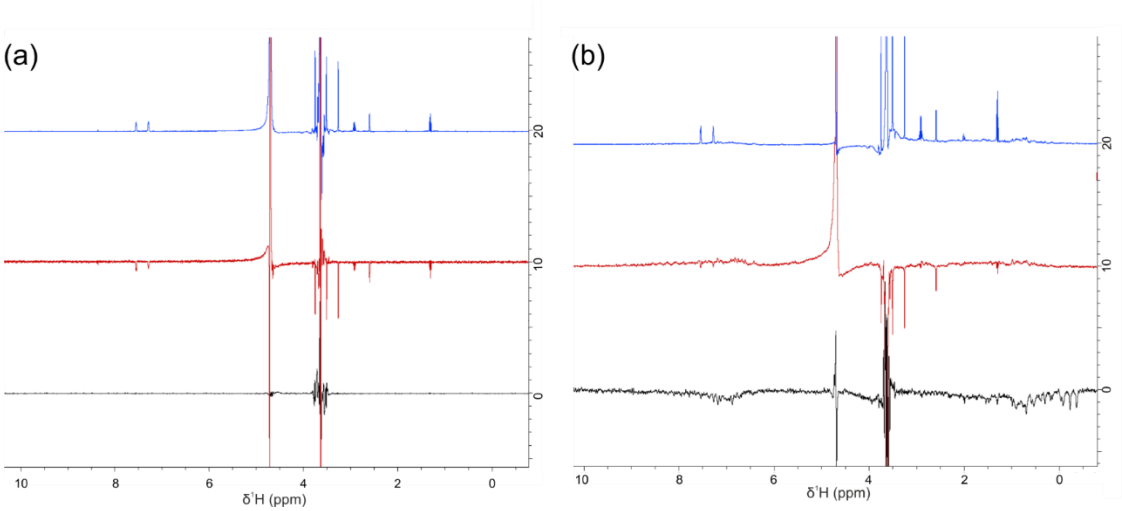


Figure A.5.6- NMR spectra collected for Compound 4.19. (a) The buffer standard, 1 mM **4.19** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.19** and 20 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

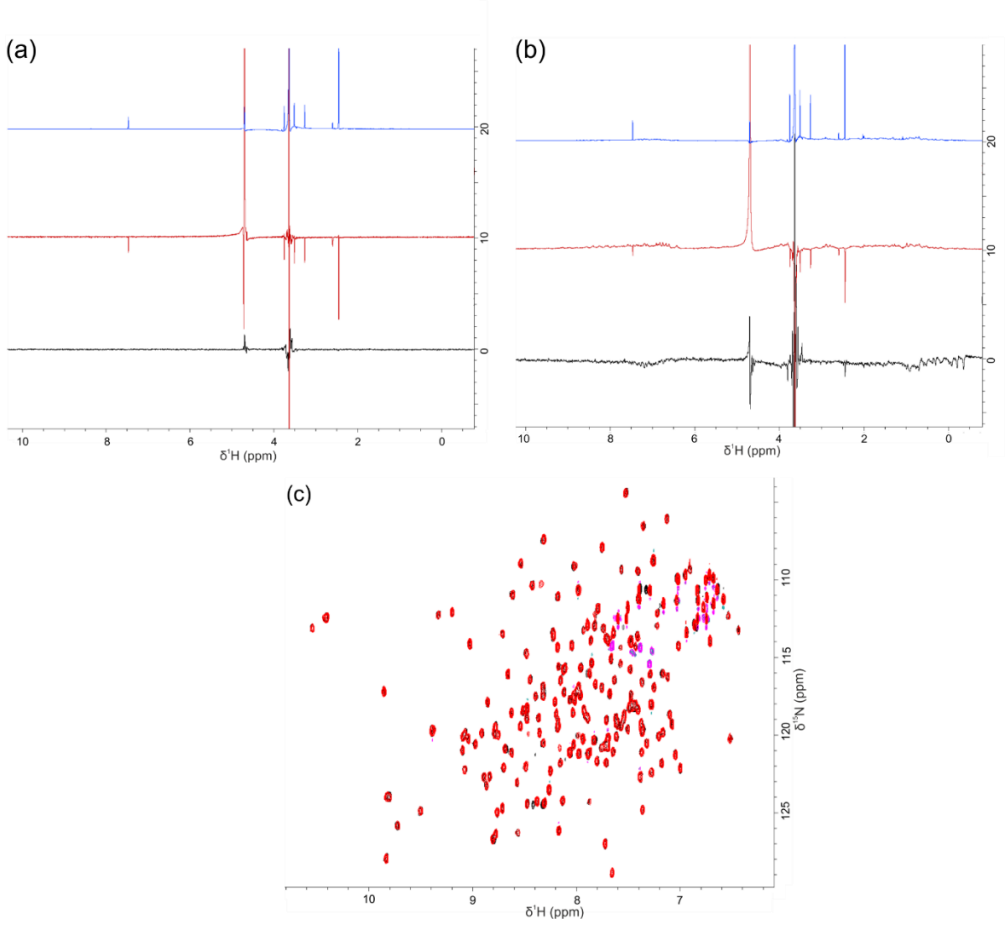


Figure A.5.7- NMR spectra collected for Compound 4.20. (a) The buffer standard, 1 mM **4.20** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.20** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.20** and NCS1 in the aliphatic region of the STD. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50µM and DMSO 1% (black) with the spectrum of the ligand 1mM in DMSO 1% and NCS1 50 µM red both suggested that this was a false result.

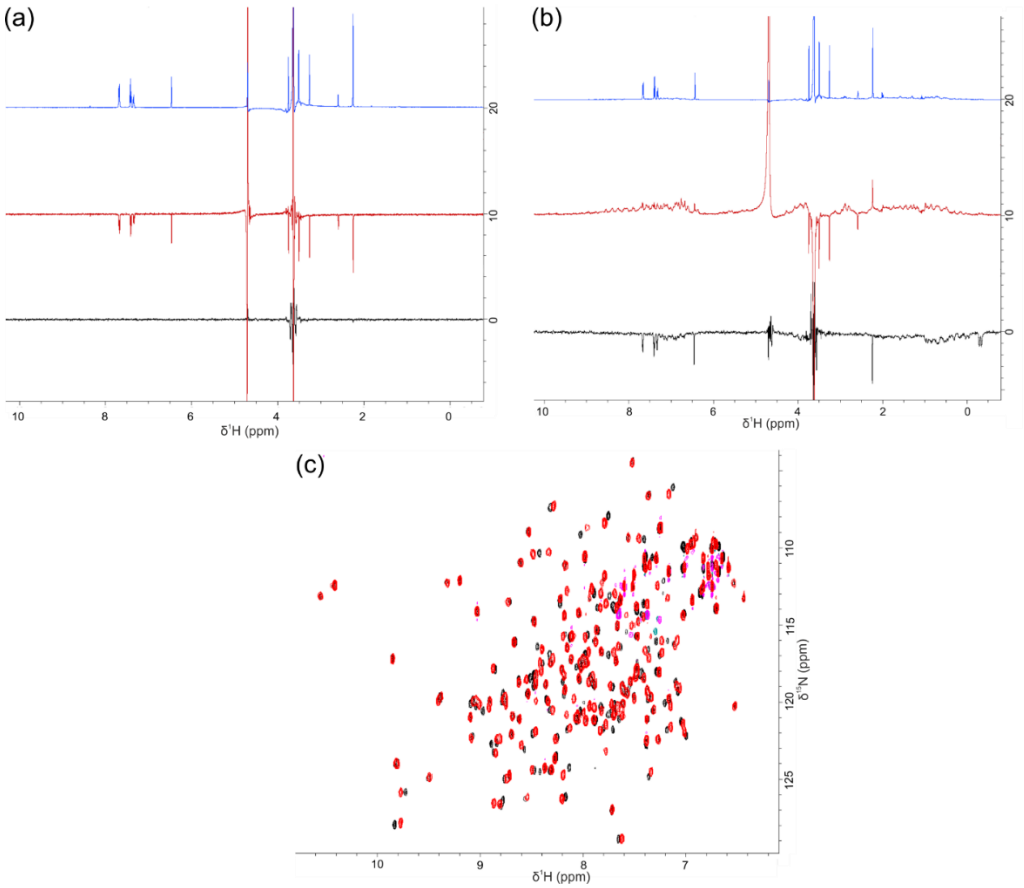


Figure A.5.8- NMR spectra collected for Compound 4.21. (a) The buffer standard, 1 mM **4.21** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.21** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible interaction between **4.21** and NCS1 can be observed in the aromatic and aliphatic regions of the STD and waterLOGSY spectra. This is observed as the ligand signals appear in the STD and their corresponding signals in the waterLOGSY are reduced or positive. An interaction between the small molecule and NCS1 was confirmed in **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red with numerous amino acid residues appearing to be perturbed by the presence of **4.21**. All three experimental results are indicative of a possible binding interaction.

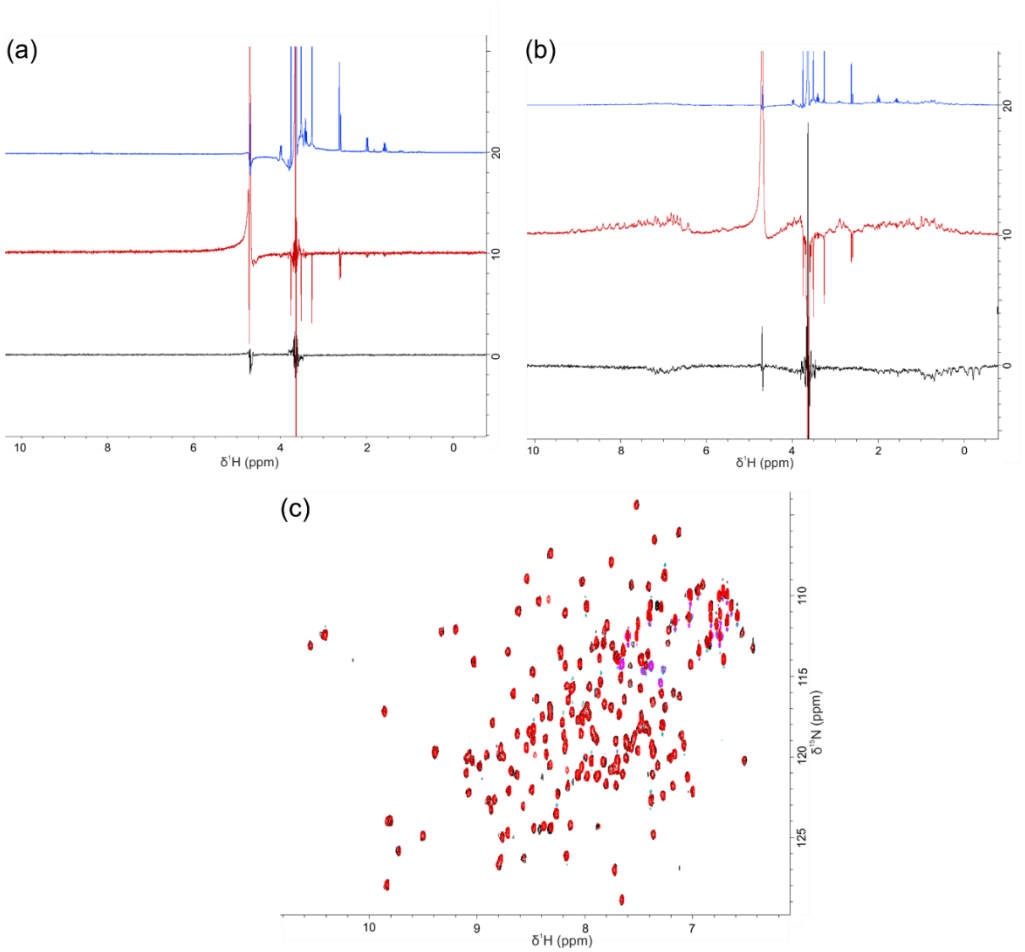


Figure A.5.9- NMR spectra collected for Compound 4.22. (a) The buffer standard, 1 mM **4.22** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.22** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect of **4.22** on NCS1 was thought to be observed in the aliphatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red both suggested that this was a false result.

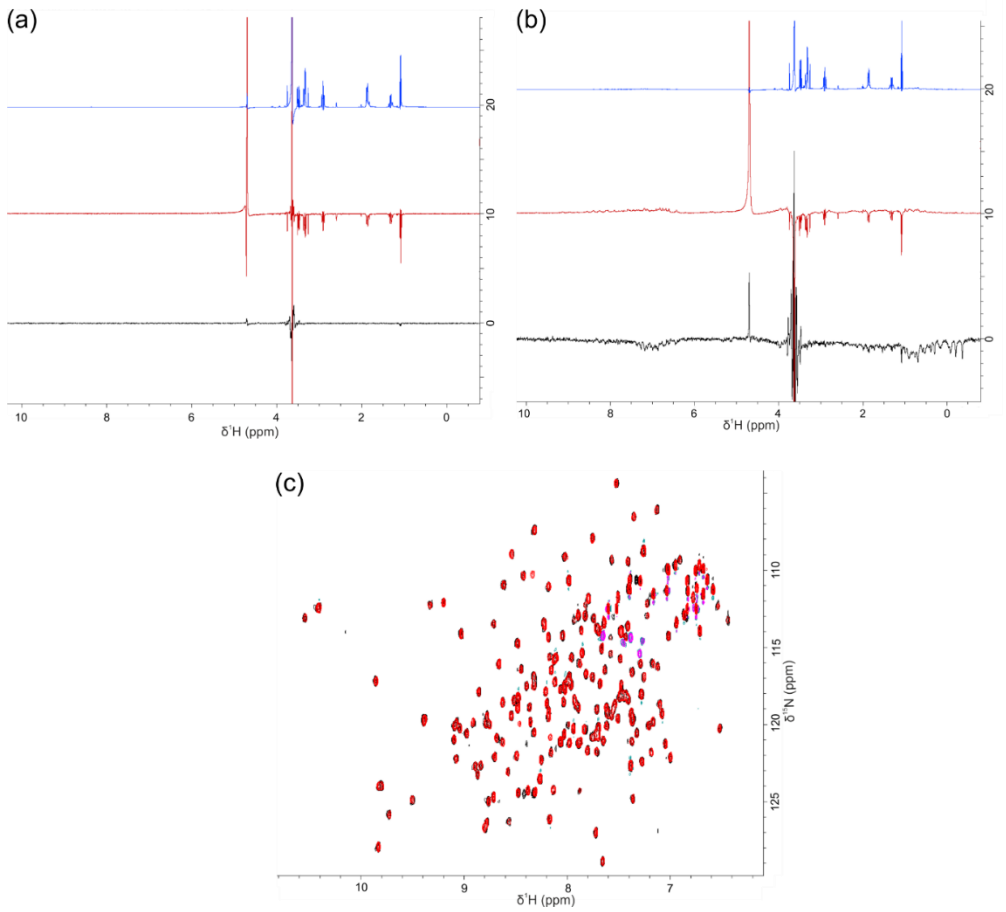


Figure A.5.11- NMR spectra collected for Compound 4.23. **(a)** The buffer standard, 1 mM **4.23** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.23** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction between **4.23** and NCS1 was thought to be observed in the aliphatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1mM in DMSO 1% and NCS1 50 µM red, both suggested that this was a false result.

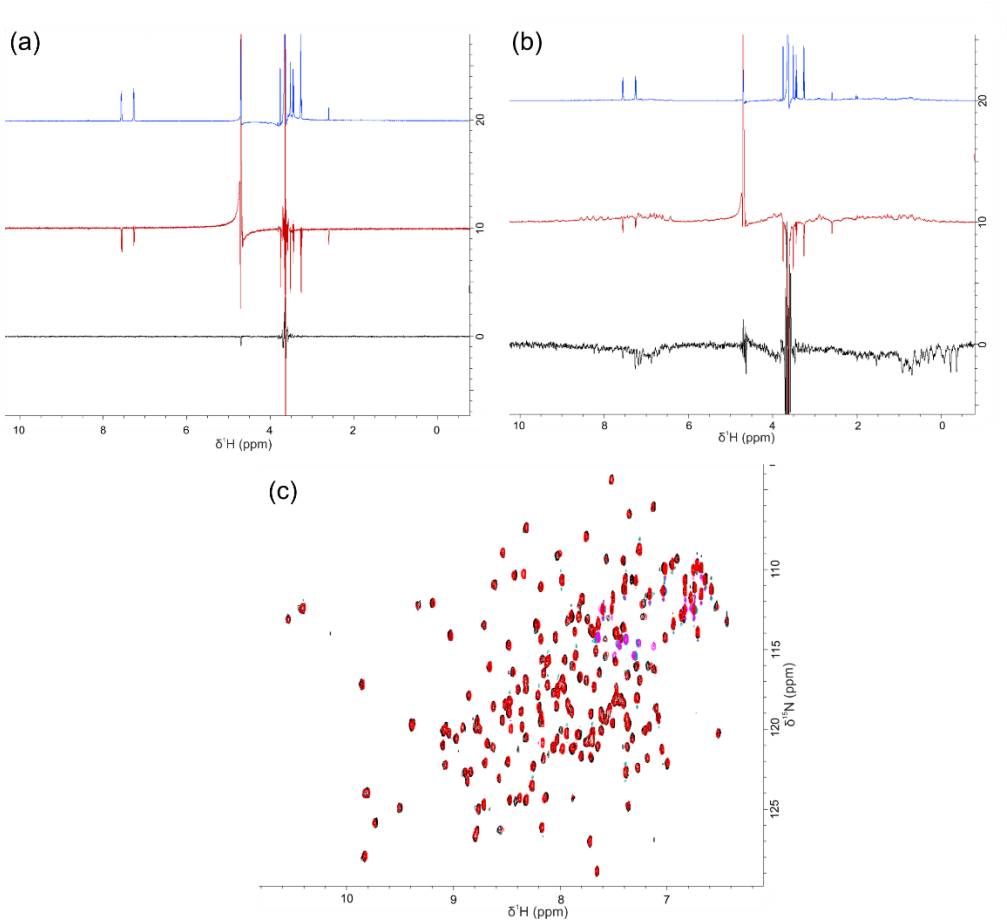


Figure A.5.12- NMR spectra collected for Compound 4.24 **(a)** The buffer standard, 1 mM **4.24** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.24** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction between the ligand NCS1 was observed in the aromatic region of the STD. This was confirmed by reduced intensity of the corresponding signals in the waterLOGSY spectrum and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, whereby some of the residues appear partially perturbed.

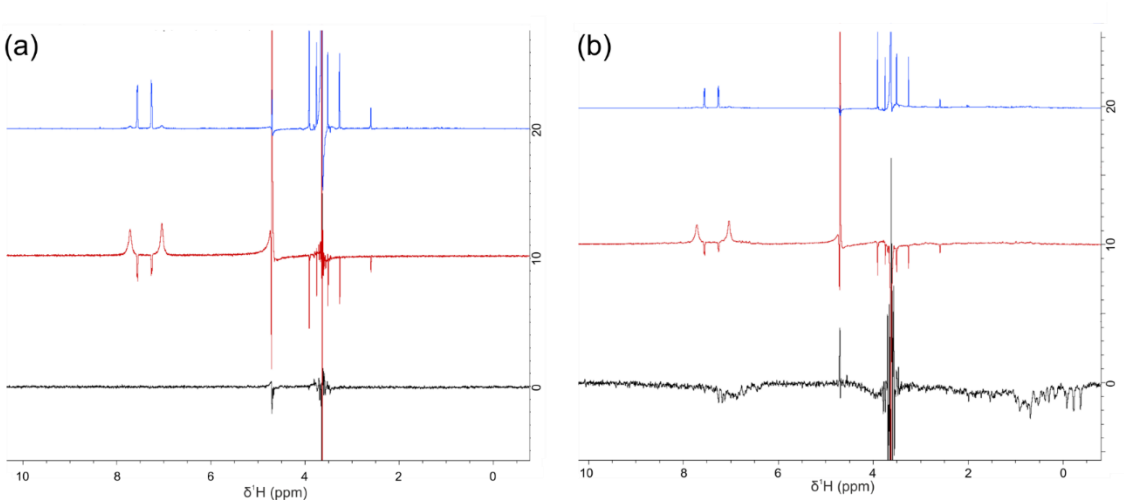


Figure A.5.13- NMR spectra collected for Compound 4.25. (a) The buffer standard, 1 mM **4.25** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.25** and 20 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

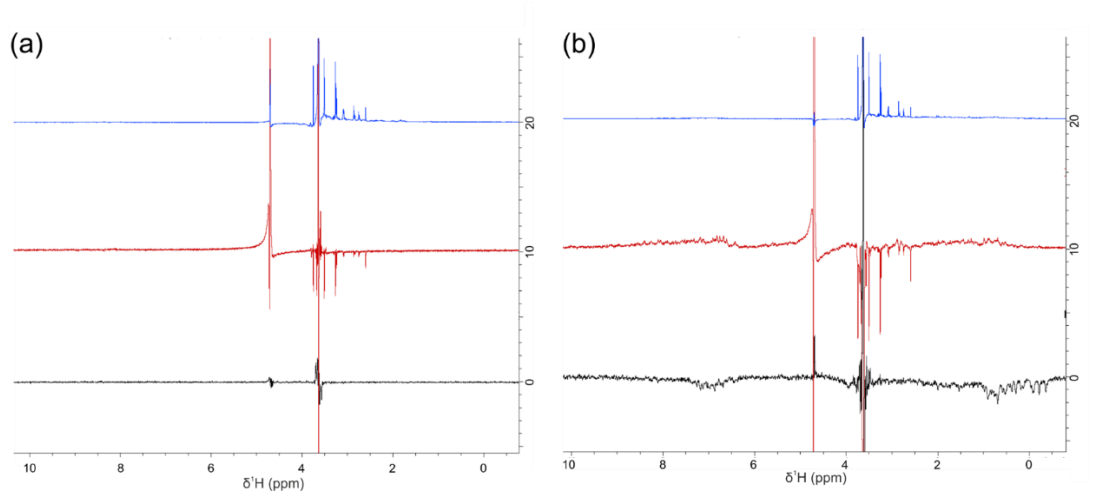


Figure A.5.14- NMR spectra collected for Compound 4.26. (a) The buffer standard, 1 mM **4.26** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.26** and 20 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

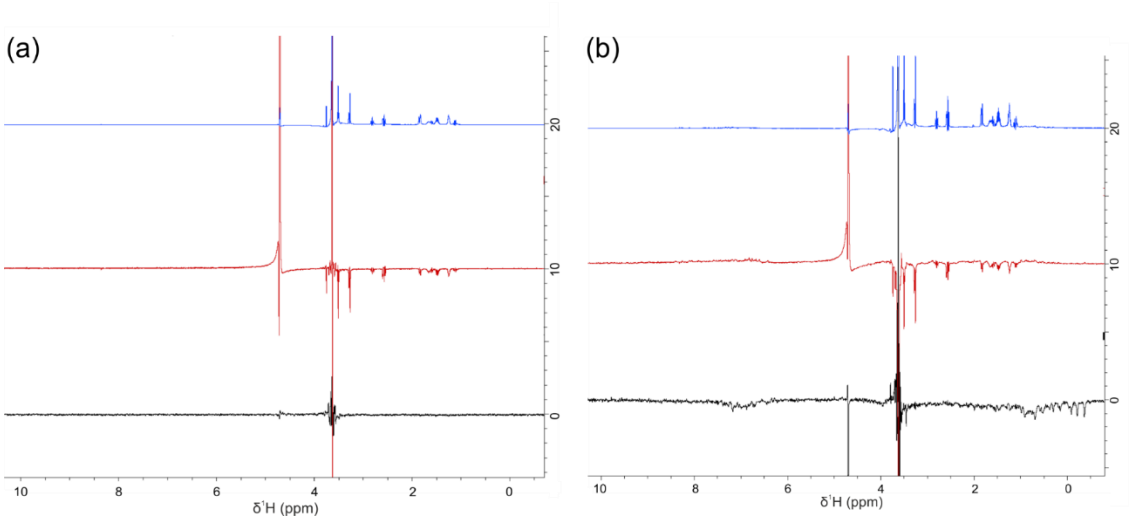


Figure A.5.15- NMR spectra collected for Compound 4.27. (a) The buffer standard, 1 mM **4.27** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.27** and 20 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an effect of the ligand on NCS1 was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

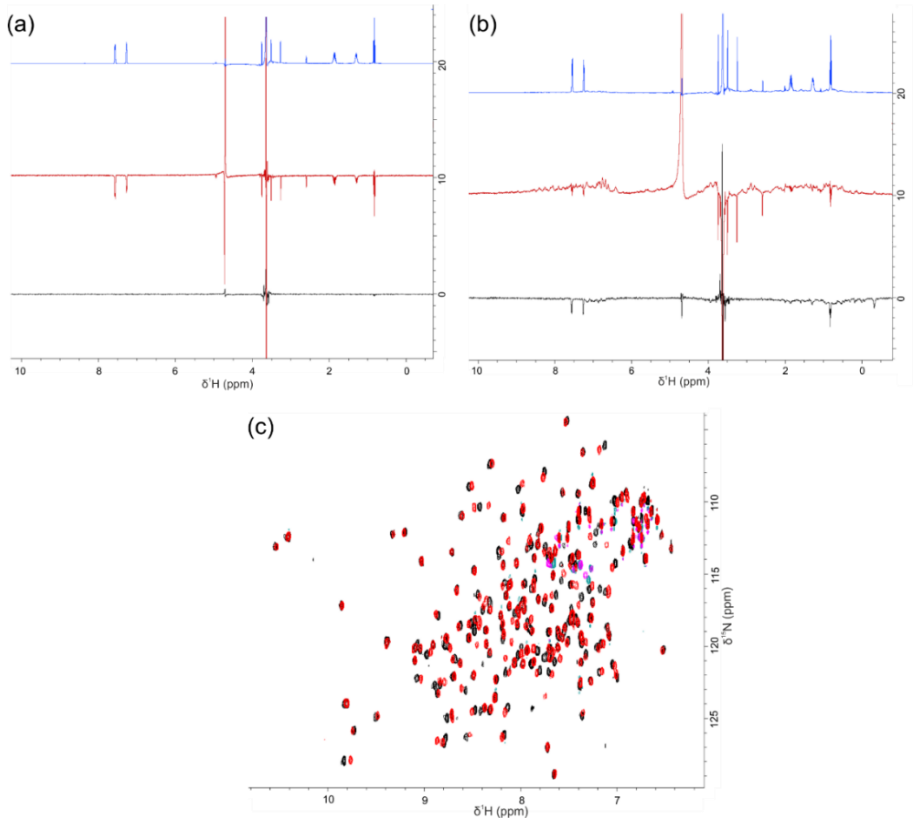


Figure A.5.16- NMR spectra collected for Compound 4.28. The buffer standard, 1 mM **4.28** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.28** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand on NCS1 can be observed in the aromatic and aliphatic regions of the STD where the ligand signals appear in the STD and waterLOGSY where their corresponding signals in the waterLOGSY are reduced or positive. An interaction between the small molecule and NCS1 was confirmed in **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red with numerous amino acid residues appearing to be perturbed by the presence of **4.28**. All three experimental results are indicative of a possible binding interaction.

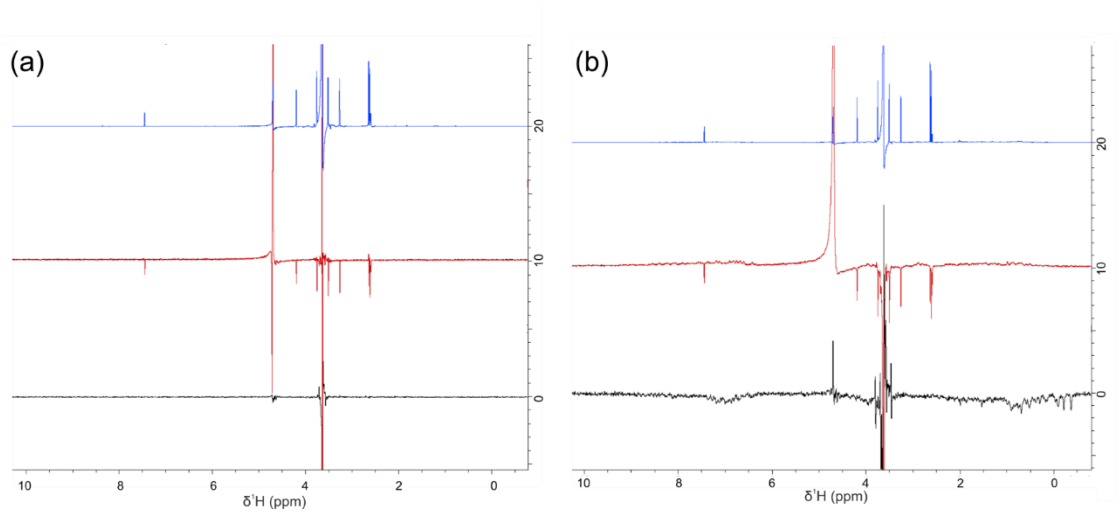


Figure A.5.17- NMR spectra collected for Compound 4.29. (a) The buffer standard, 1 mM **4.29** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. **(b)** 1 mM **4.29** and 20 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum seen in red and STD experiment black. No indication of an interaction was observed in these experiments and so a ^{1}H ^{15}N HSQC spectra was not collected.

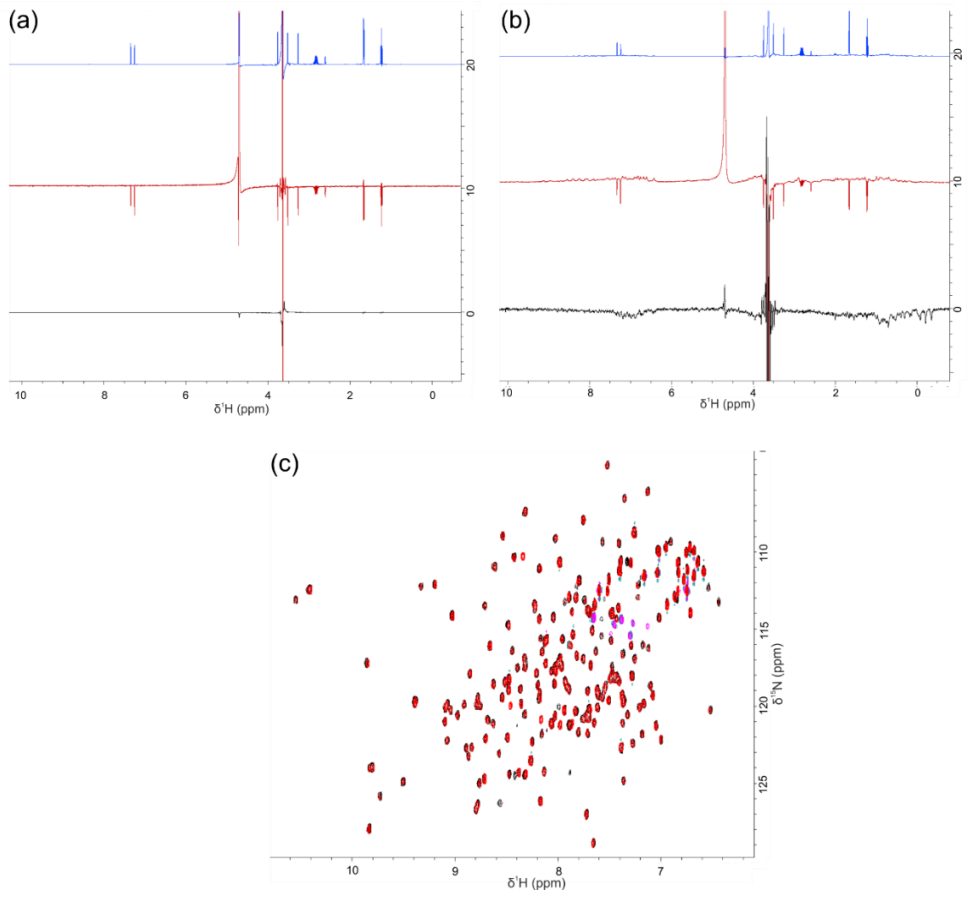


Figure A.5.18- NMR spectra collected for Compound 4.30. (a) The buffer standard, 1 mM **4.30** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.30** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect was thought to be observed in the aliphatic region of the STD. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, both suggested that this was a false result.

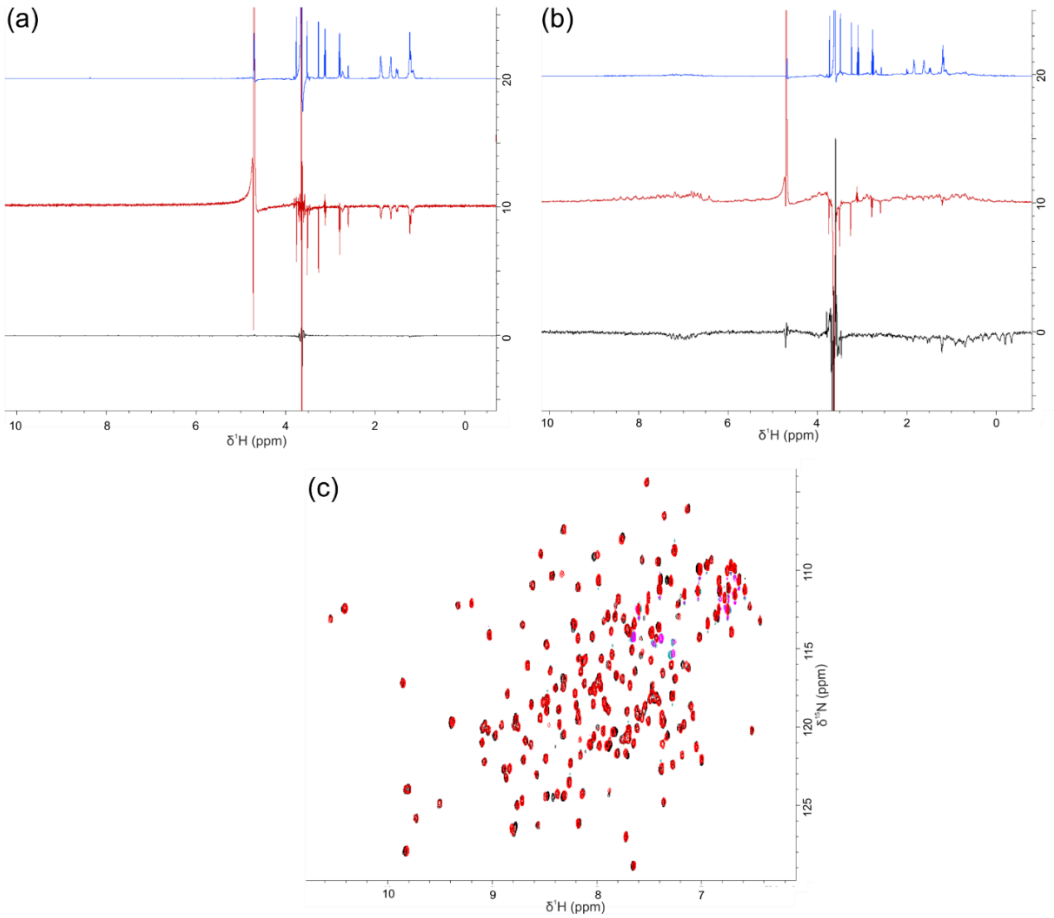


Figure A.5.19- NMR spectra collected for Compound 4.31. (a) The buffer standard, 1 mM **4.31** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.31** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect was observed up-field in the aliphatic region of the STD. This was confirmed by reduced intensity of the corresponding signals in the waterLOGSY spectrum and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, where some of the residues appear partially perturbed.

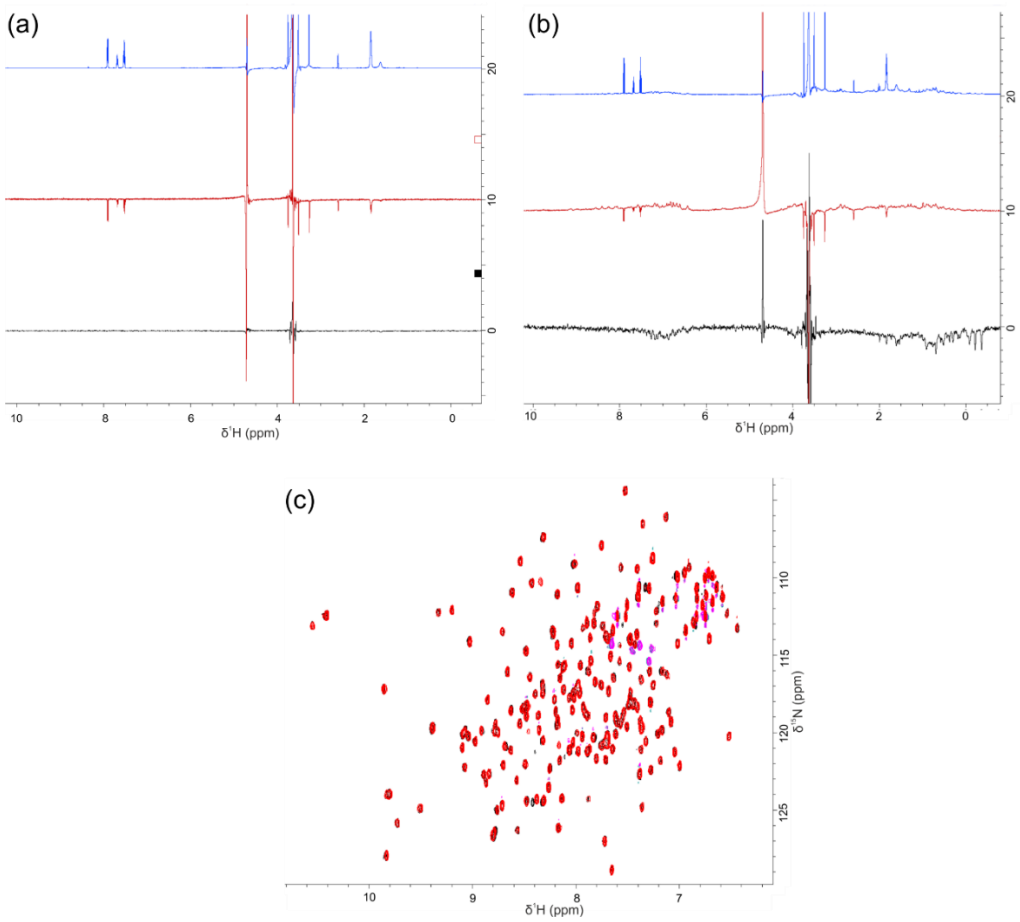


Figure A.5.20- NMR spectra collected for Compound 4.32. **(a)** The buffer standard, 1 mM **4.32** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.32** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect was observed up-field in the aliphatic region of the STD. This was confirmed by reduced intensity of the corresponding signals in the waterLOGSY spectrum and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, where some of the residues appear partially perturbed.

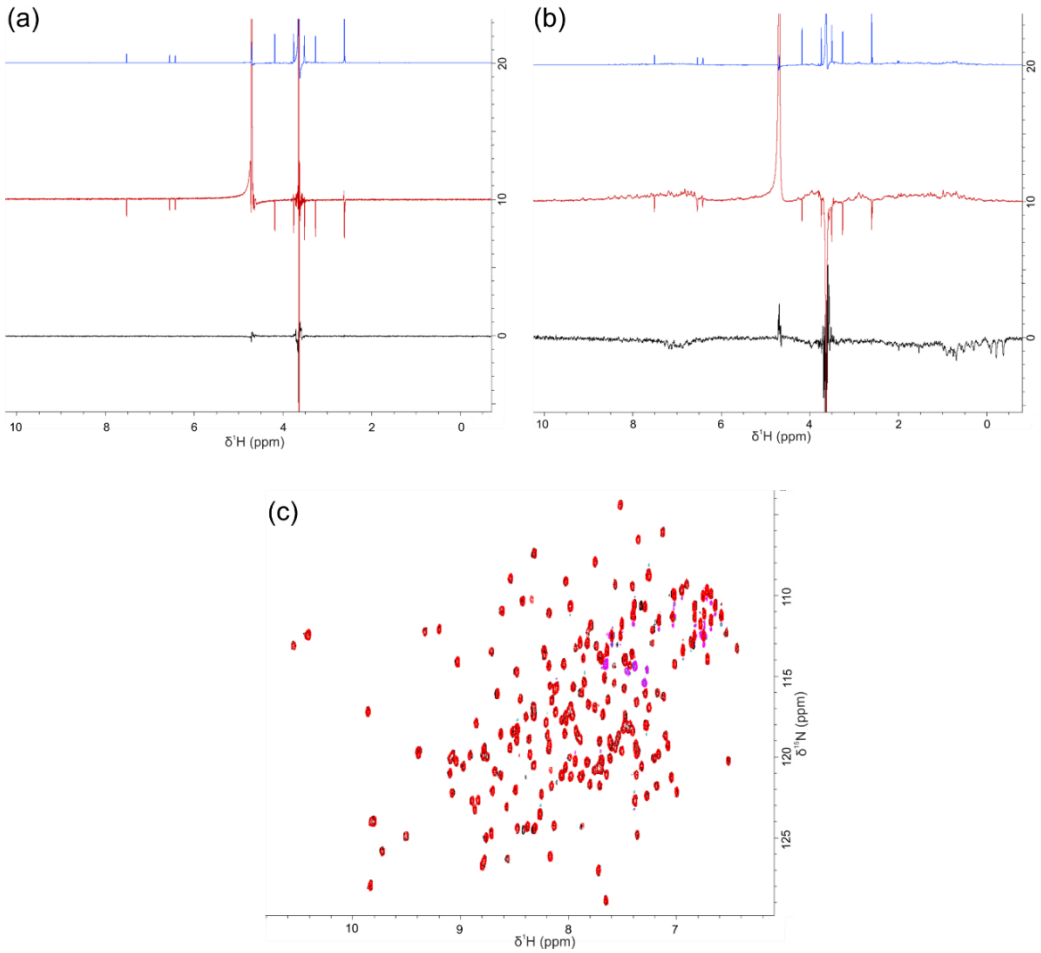


Figure A.5.21- NMR spectra collected for Compound 4.33. **(a)** The buffer standard, 1 mM **4.33** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.33** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect between **4.33** and NCS1 was observed up-field in the aliphatic region of the STD. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, both suggested that this was a false result.

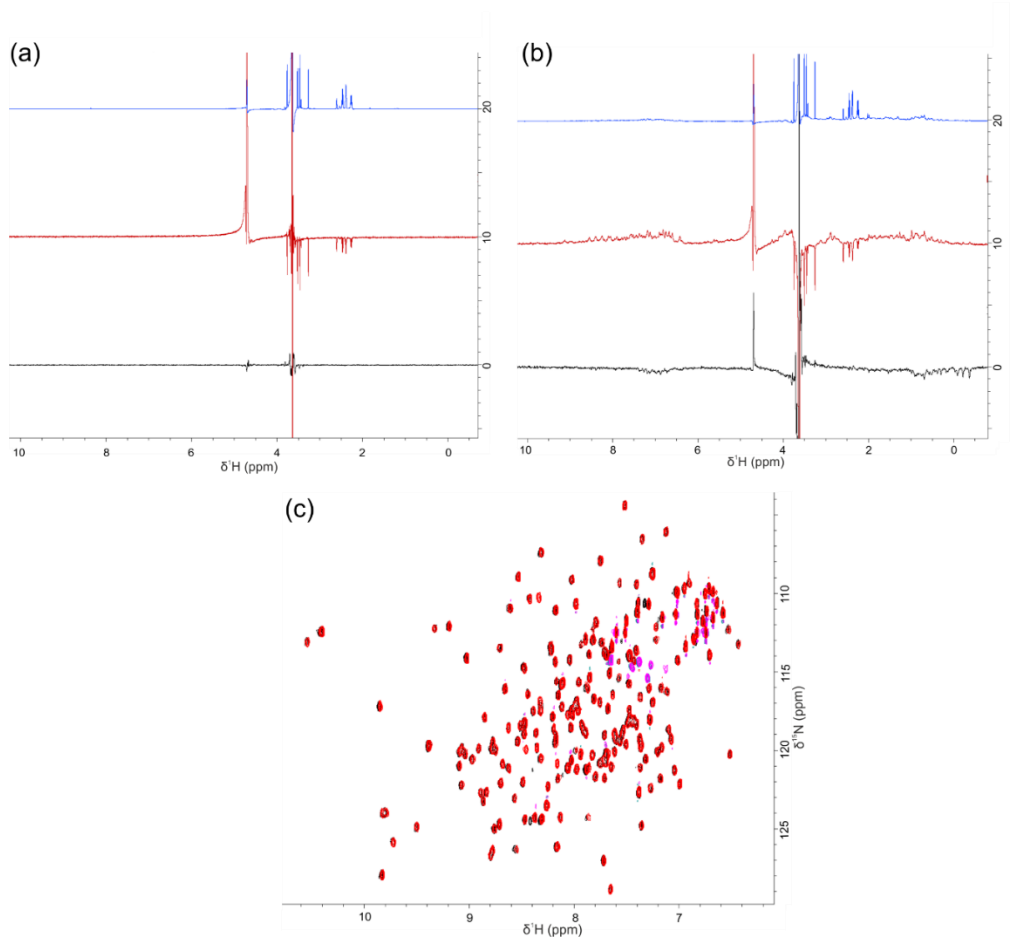


Figure A.5.22- NMR spectra collected for Compound 4.34. **(a)** The buffer standard, 1 mM **4.34** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.34** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect between **4.34** and NCS1 was observed up-field in the aliphatic region of the STD. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 0.95% (black) with the spectrum of the ligand 1 mM in DMSO 0.95% and NCS1 50µM red, both suggested that this was a false result.

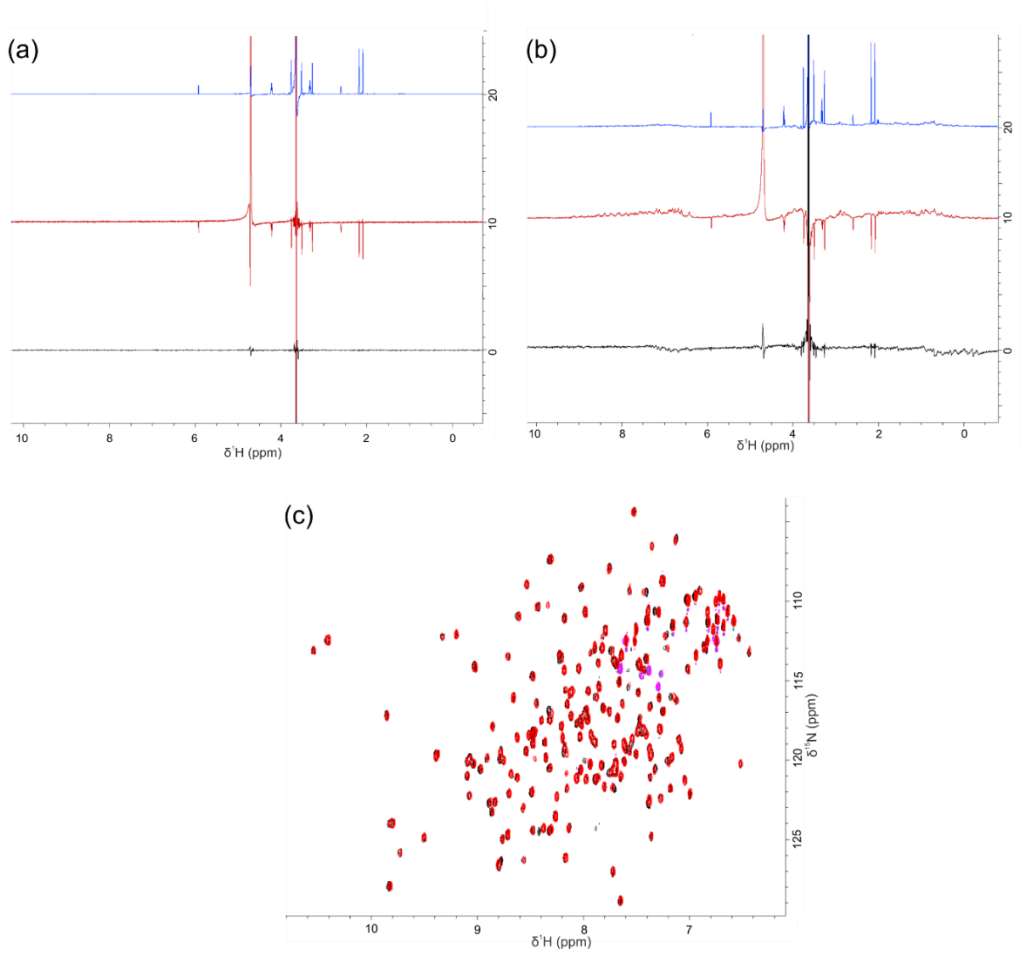


Figure A.5.23- NMR spectra collected for Compound 4.35 **(a)** The buffer standard, 1 mM **4.35** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.35** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect between **4.35** and NCS1 was observed up-field in the aliphatic region of the STD. However the waterLOGSY spectra and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, both suggested that this was a false result.

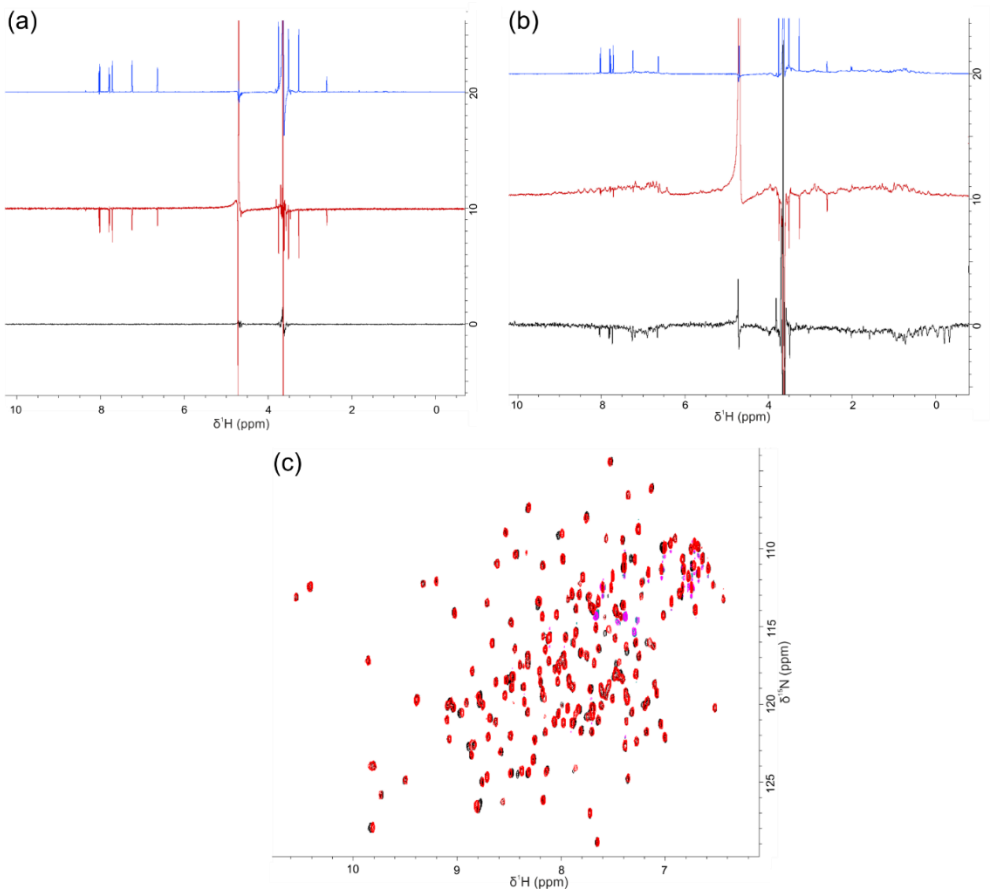


Figure A.5.24- NMR spectra collected for Compound 4.36. **(a)** The buffer standard, 1 mM **4.36** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.36** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand upon NCS1 can be observed in the aromatic region of the STD spectra where the ligand signals are apparent and waterLOGSY experiment where the corresponding signals are reduced or positive. An interaction between the small molecule and NCS1 was confirmed in **(c)** an overlay of two 2-D HSQC spectra. The control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, with the chemical shift of some amino acid residues appearing to be perturbed by the presence of **4.36**. All three experimental results are indicative of a possible binding interaction.

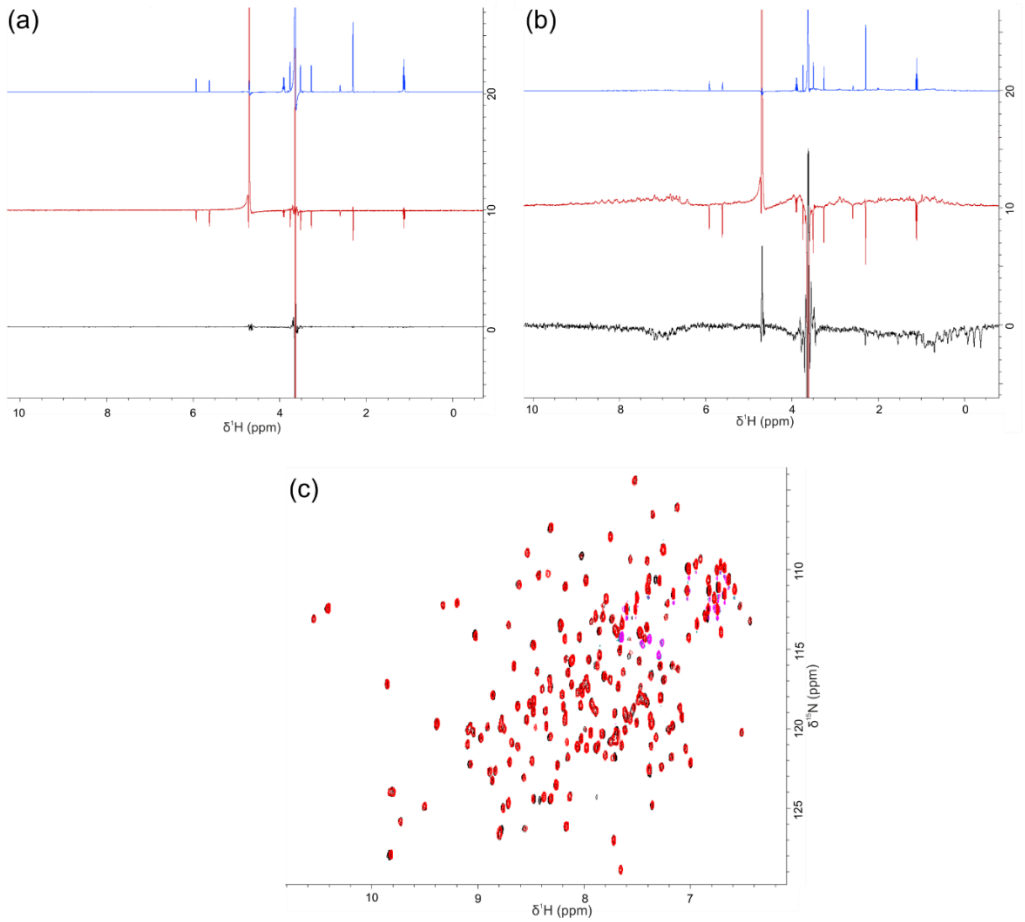


Figure A.5.25- NMR spectra collected for Compound 4.37. (a) The buffer standard, 1 mM **4.37** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.37** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.37** and NCS1, in the aliphatic and aromatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, both suggested that this was a false result.

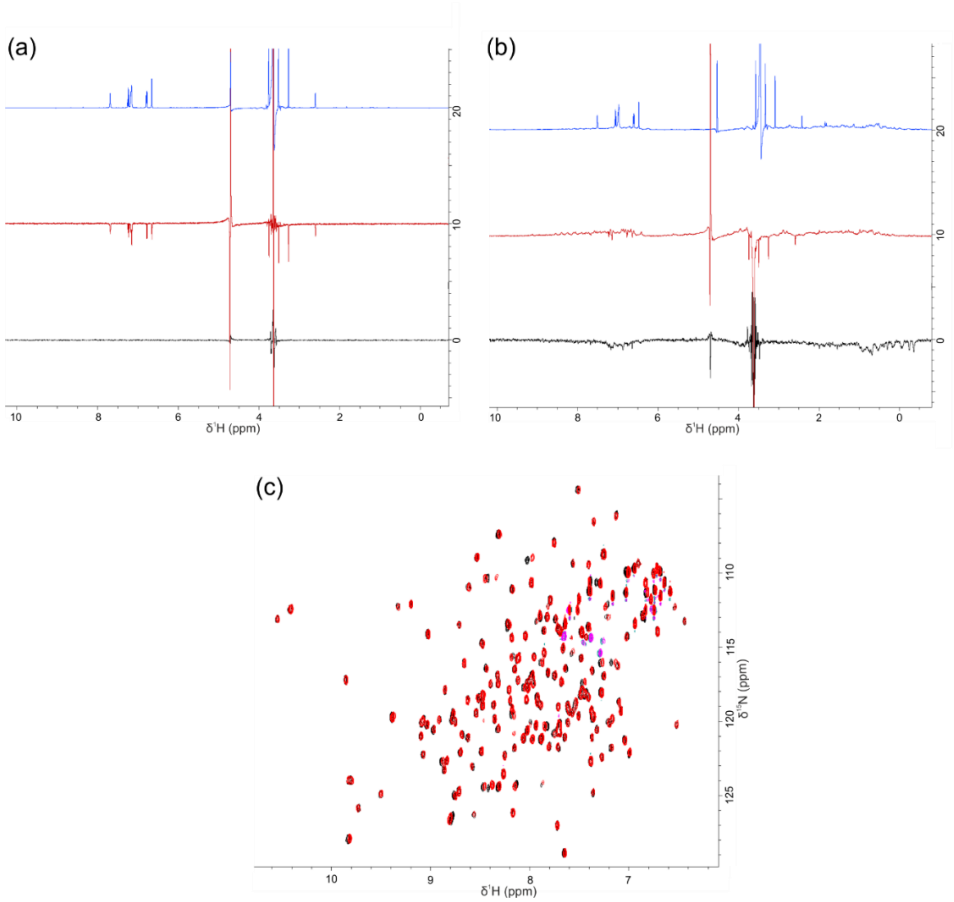


Figure A.5.26- NMR spectra collected for Compound 4.38. (a) The buffer standard, 1 mM **4.38** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.38** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.38** and NCS1, observable in the aromatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, both suggest that this was a false result.

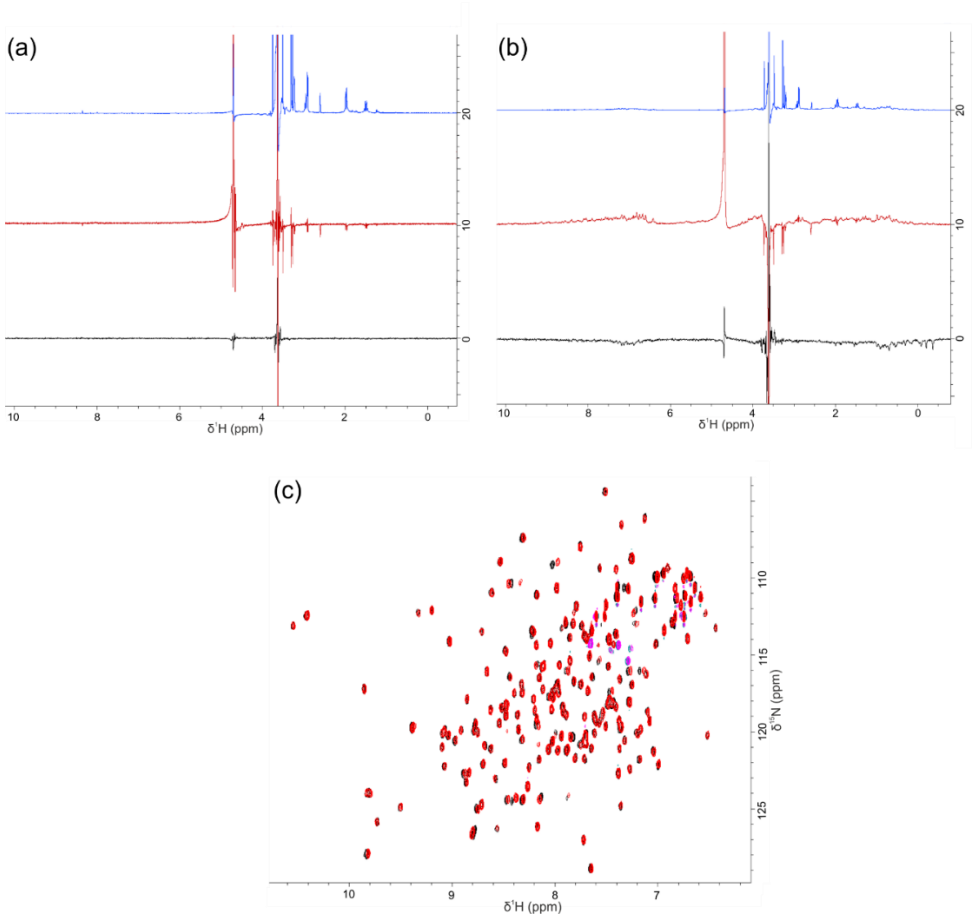


Figure A.5.27- NMR spectra collected for Compound 4.39. (a) The buffer standard, 1 mM **4.39** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.39** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.39** and NCS1, in the aliphatic region of the STD spectra. However the waterLOGSY experiment and **(c)** an overlay of two 2-D HSQC spectra, the control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 μM red, both suggested that this was a false result.

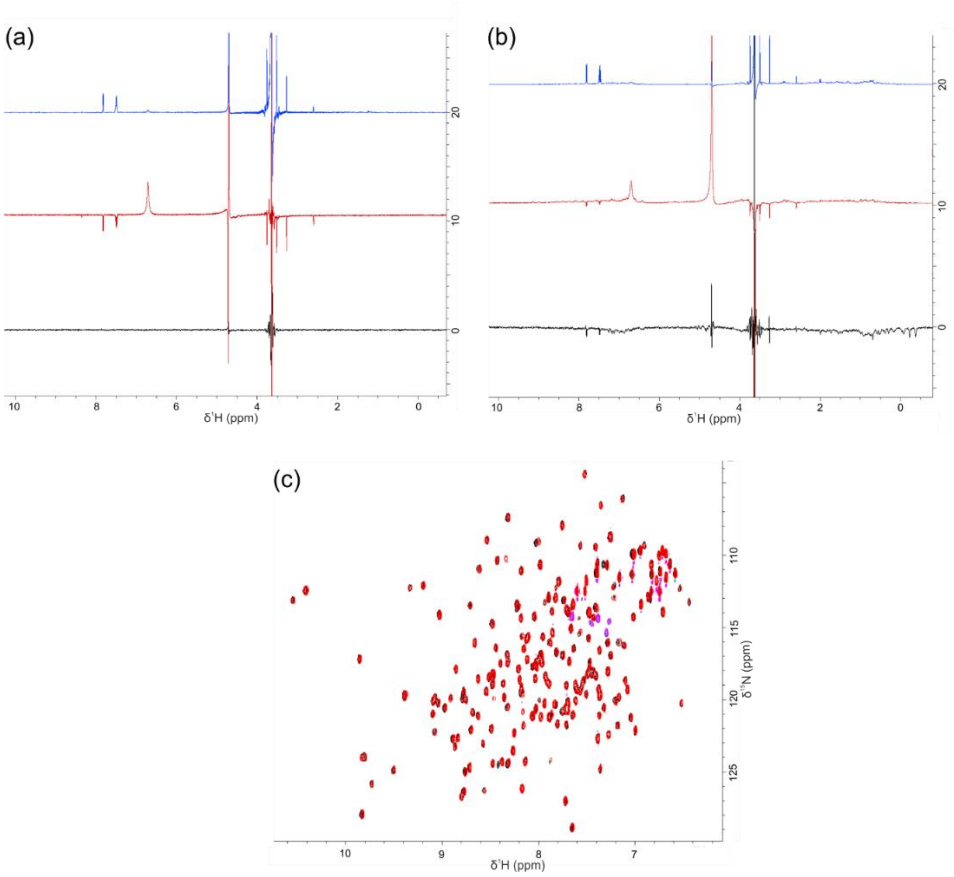


Figure 8.5.28- **NMR spectra collected for Compound 4.40** **(a)** The buffer standard, 1 mM **4.40** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.40** and 50 µM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand upon NCS1 can be observed in the down-field aromatic region of the STD spectra where some of the ligand signals are apparent and waterLOGSY experiment where the corresponding signals are reduced or positive. An interaction between the small molecule and NCS1 was confirmed in **(c)** an overlay of two 2-D HSQC spectra. The control spectra of NCS1 50 µM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50 µM red, with the chemical shift of some amino acid residues appearing to be perturbed by the presence of **4.40**. All three experimental results are indicative of a possible binding interaction.

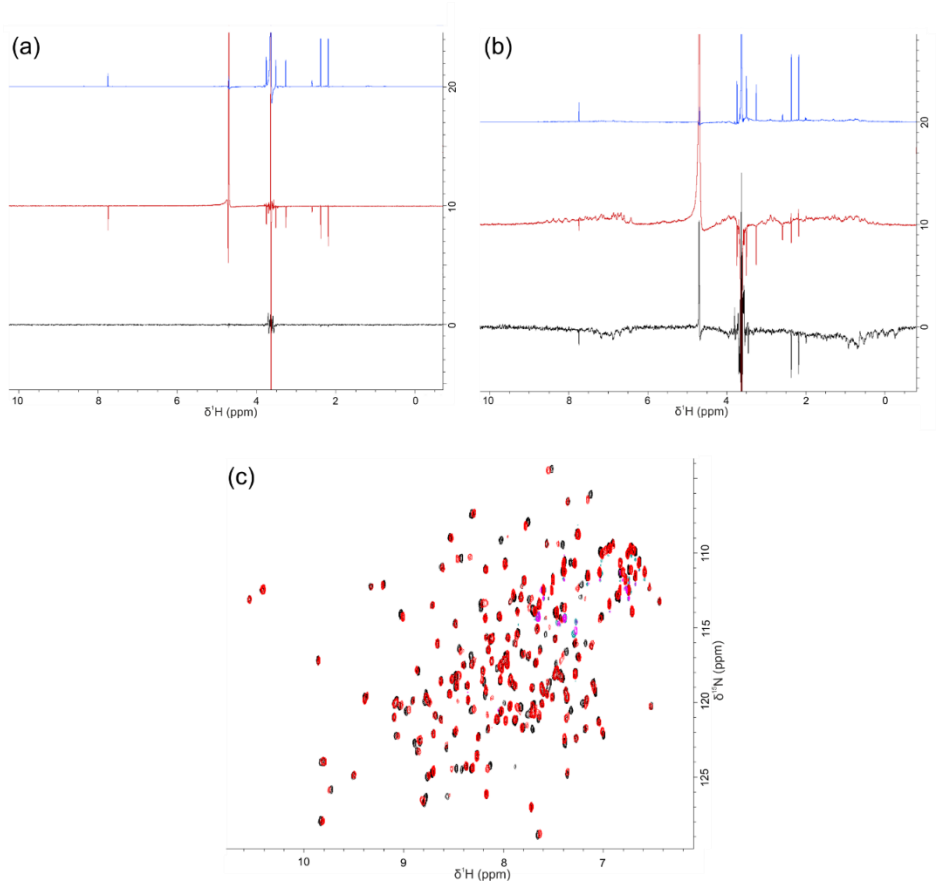


Figure A.5.29- **NMR spectra collected for Compound 4.41**. **(a)** The buffer standard, 1 mM **4.41** in NMR buffer with the 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** 1 mM **4.41** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand upon NCS1 can be observed in the up-field aliphatic region of the STD spectra where some of the ligand signals are apparent and waterLOGSY experiment where the corresponding signals are reduced or positive. An interaction between the small molecule and NCS1 was confirmed in **(c)** an overlay of two 2-D HSQC spectra. The control spectra of NCS1 50 μM and DMSO 1% (black) with the spectrum of the ligand 1 mM in DMSO 1% and NCS1 50μM red, with the chemical shift of some amino acid residues appearing to be perturbed by the presence of **4.41**. All three experimental results are indicative of a possible binding interaction.

A.6 D2R Competition experiments

To determine if the hit binding compounds of the small molecule library were allosteric or orthosteric in nature, competition experiments were designed using the peptide D2R. However as seen in the D2R peptide standard (Figure A.6.1), the peptide is poorly soluble at the desired concentration and so may not be at the required concentration to out compete the ligands.

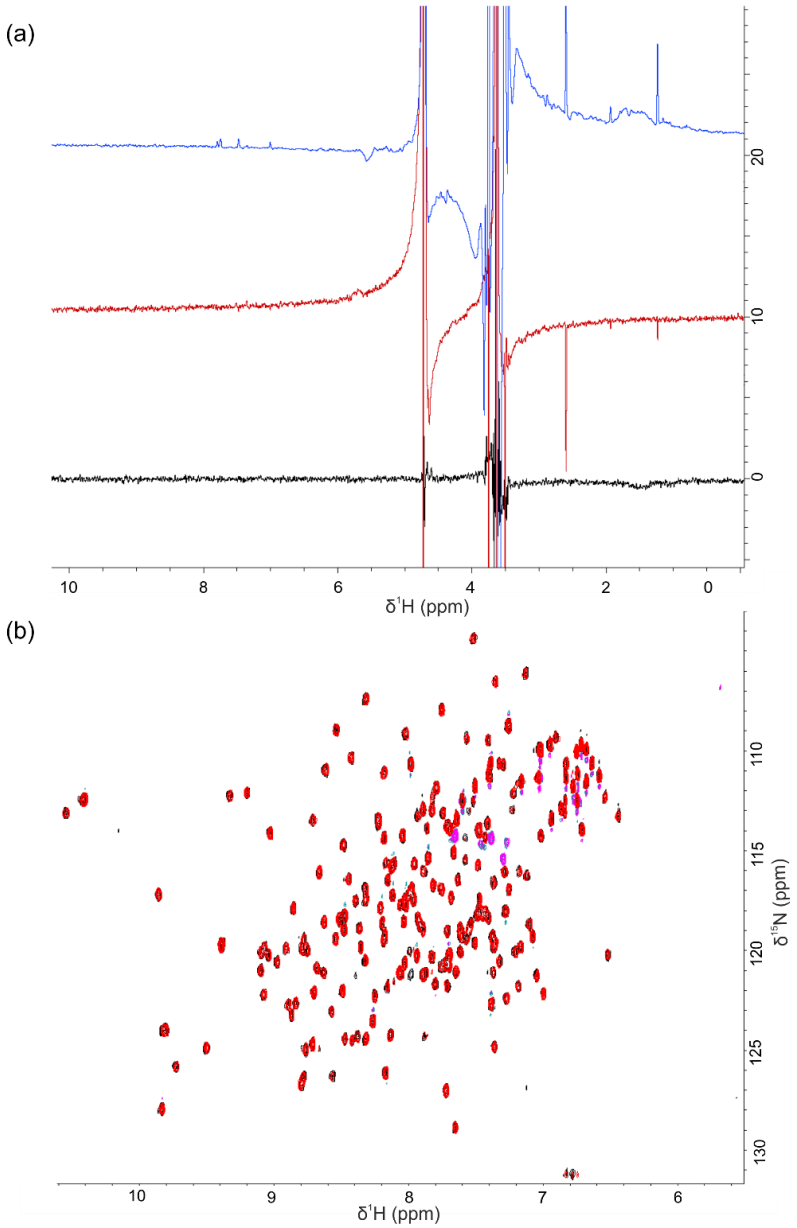


Figure A.6.1- D2R peptide standard. (a) 50 µM D2R peptide in NMR buffer, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. **(b)** Two 2-D ^{1}H ^{15}N HSQC spectra. The control spectra of NCS1 50 µM in NMR buffer (black) with the spectrum of the peptide 50 µM NCS1 50 µM red with no apparent changes in the residues observed.

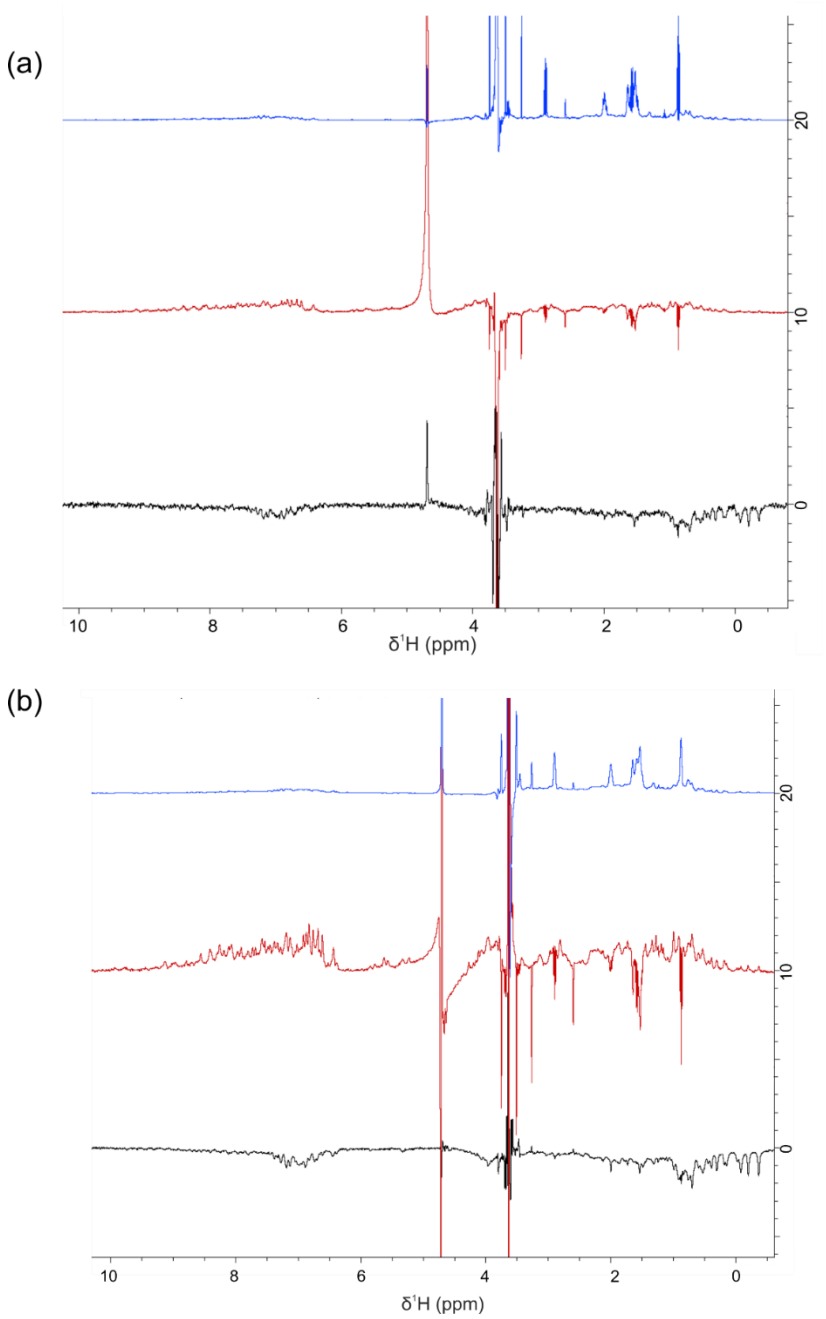


Figure A.6.2– 4.16, NCS1and D2R peptide. (a) 1 mM **4.16** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to be observed in the aliphatic region of the STD spectra **(b)** 1 mM **4.16** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

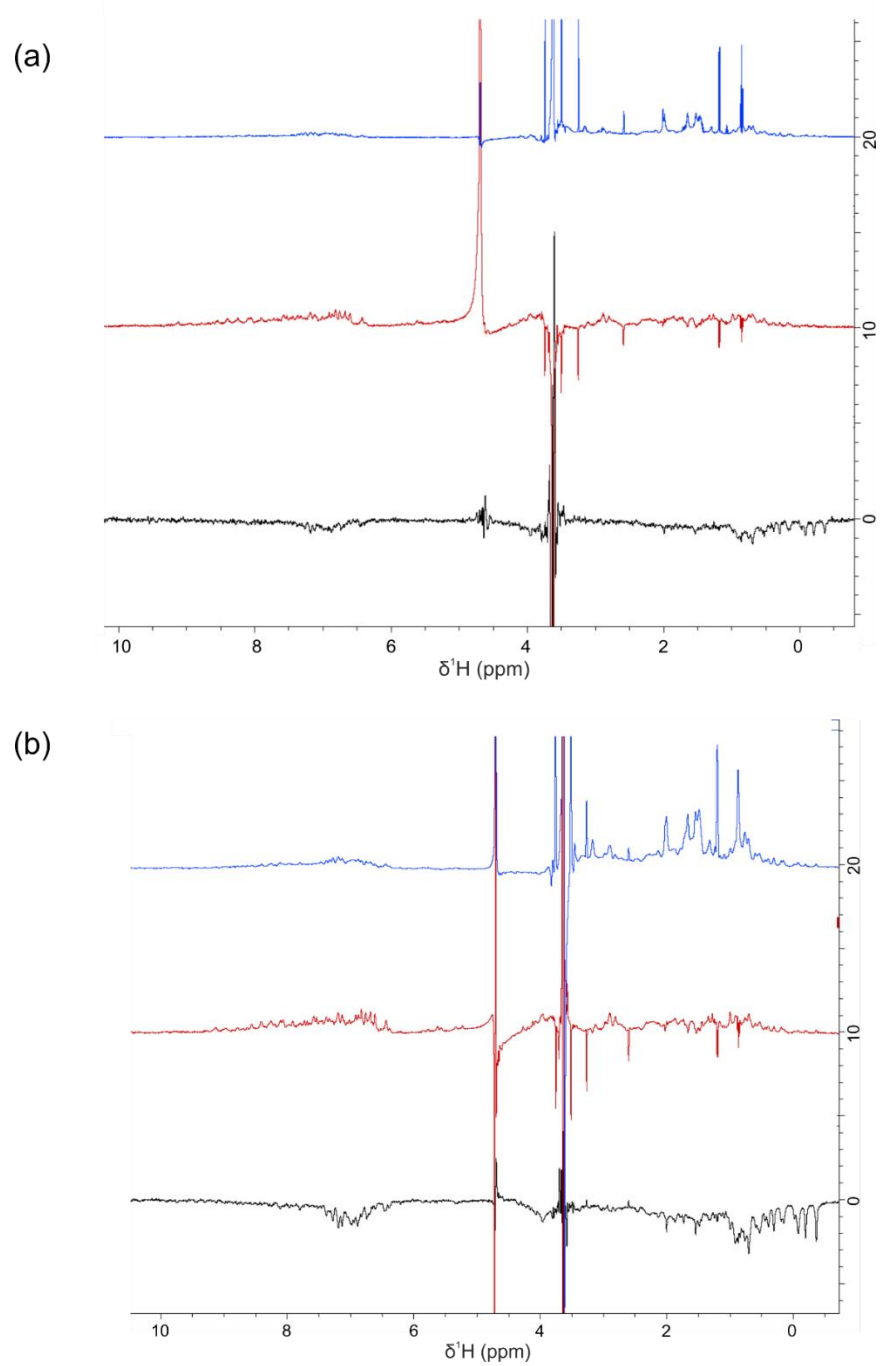


Figure A.6.3– 4.17, NCS1 and D2R peptide. (a) 1 mM **4.17** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction was thought to occur between **4.17** and NCS1, in the aliphatic region of the STD spectra. **(b)** 1 mM **4.17** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

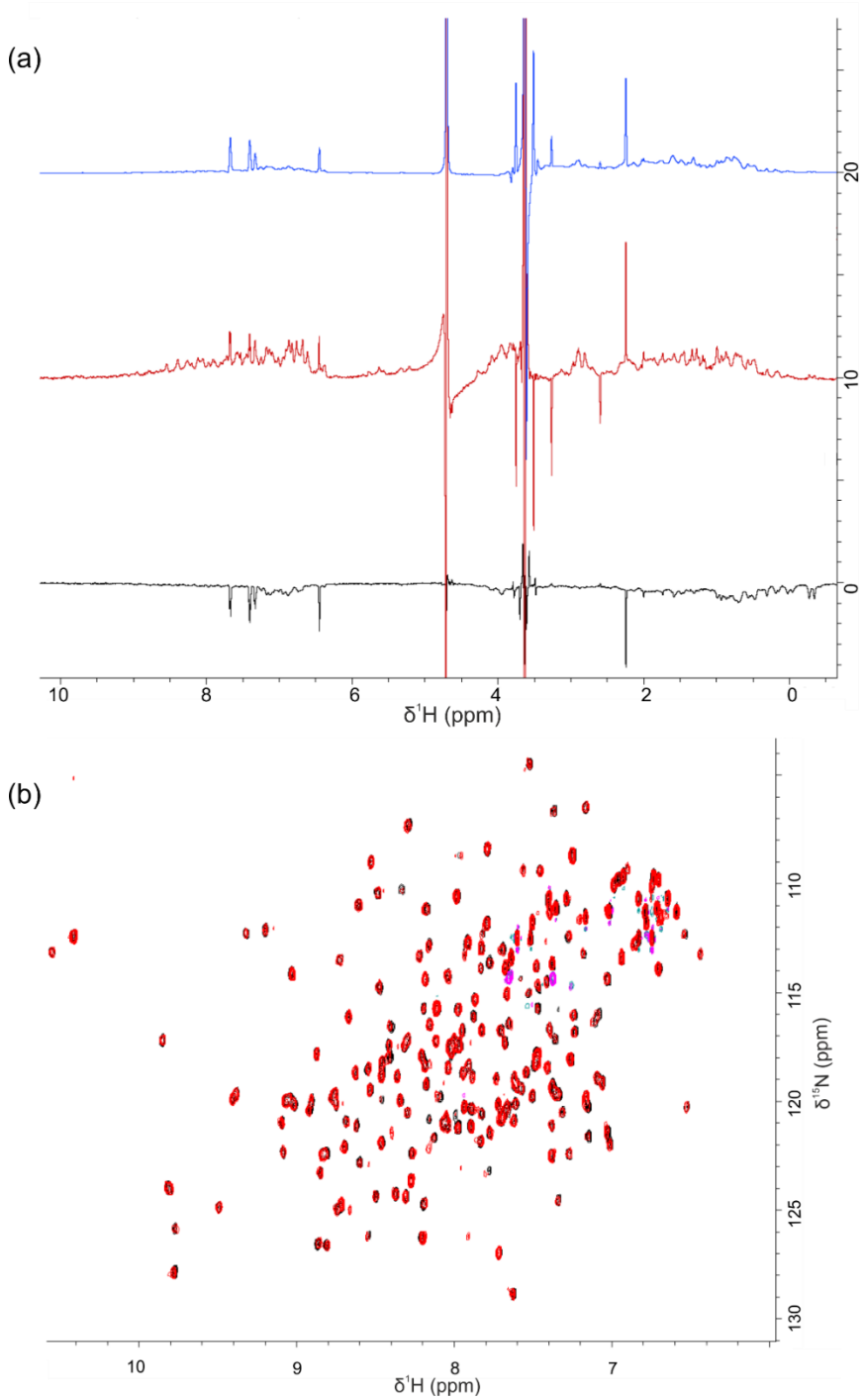


Figure A.6.4– 4.21, NCS1 and D2R peptide. (a) Two 2-D ^{1}H ^{15}N HSQC spectra. The control spectra of 1 mM C9, NCS1 50 μM in NMR buffer (black) with the spectrum of the **4.21** 50 μM, peptide 50 μM and NCS1 50 μM red with no apparent changes in the residues observed. **(b)** 1 mM **4.21** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black.

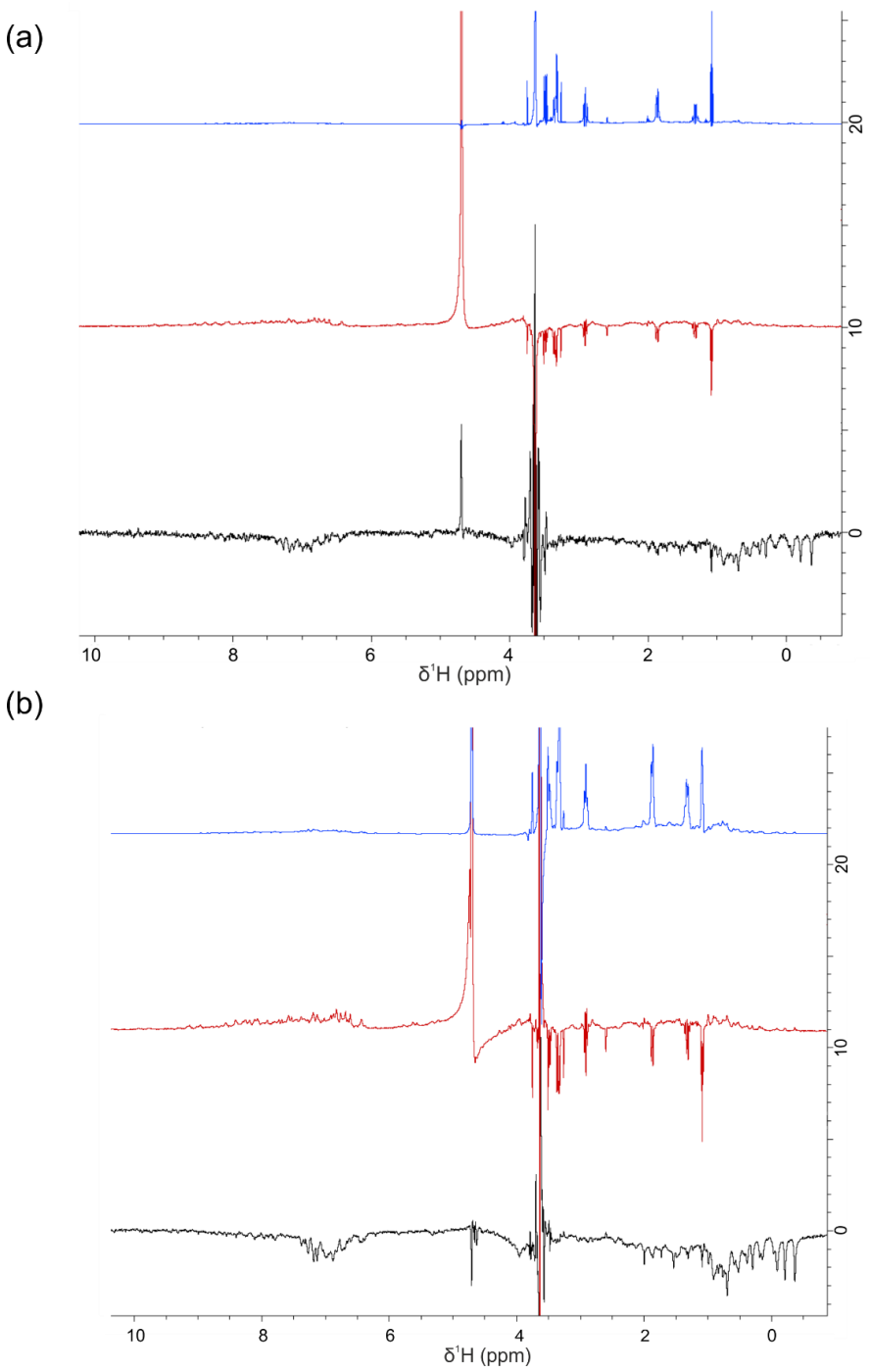


Figure A.6.5- 4.23, NCS1 and D2R peptide. (a) 1 mM **4.23** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction between 4.23 and NCS1 was thought to be observed in the aliphatic region of the STD spectra **(b)** 1 mM **4.23** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

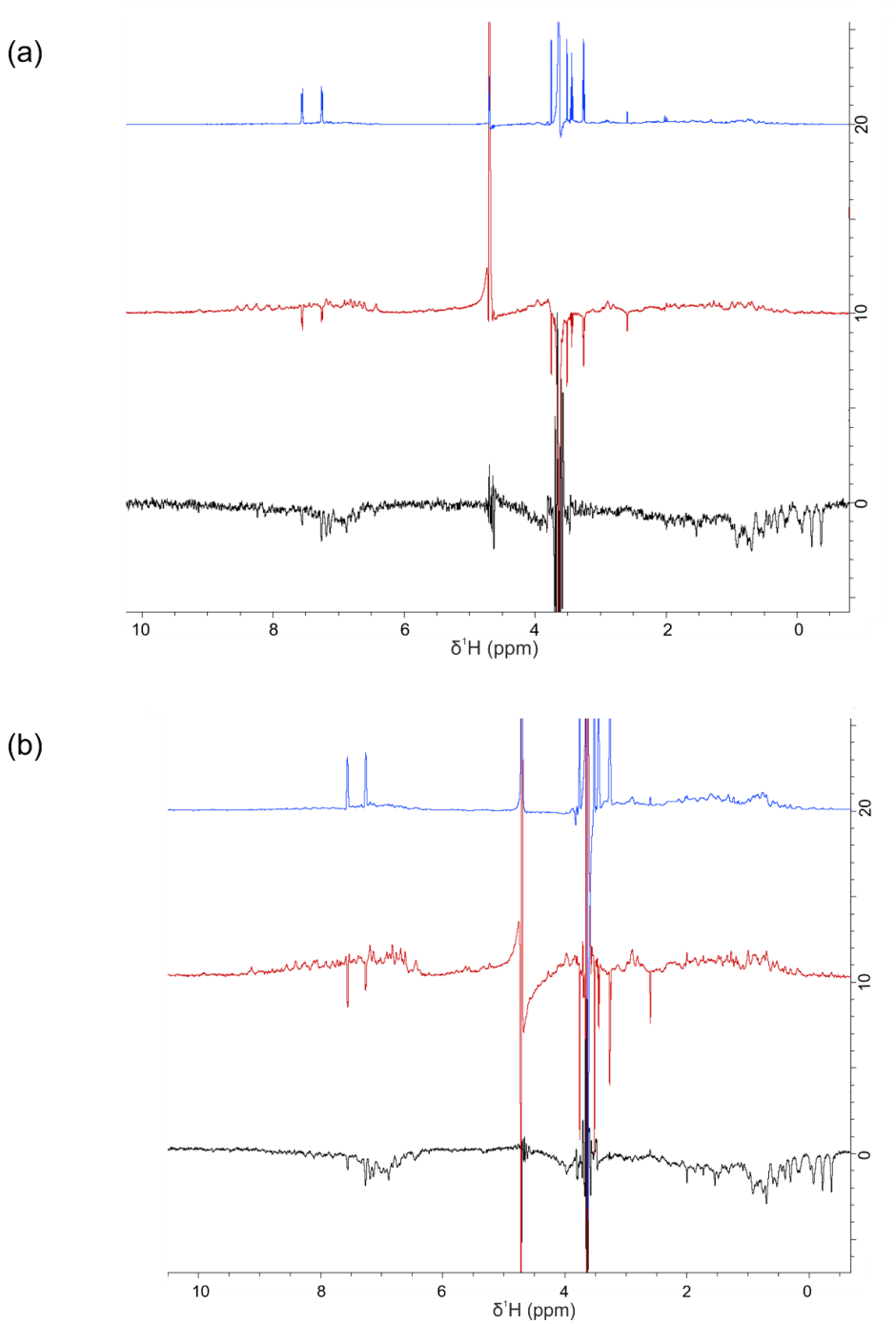


Figure A.6.6- 4.24, NCS1 and D2R peptide. (a) 1 mM **4.24** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak interaction between the ligand NCS1 was observed in the aromatic region of the STD. This was confirmed by reduced intensity of the corresponding signals in the waterLOGSY spectrum. **(b)** 1 mM **4.24** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

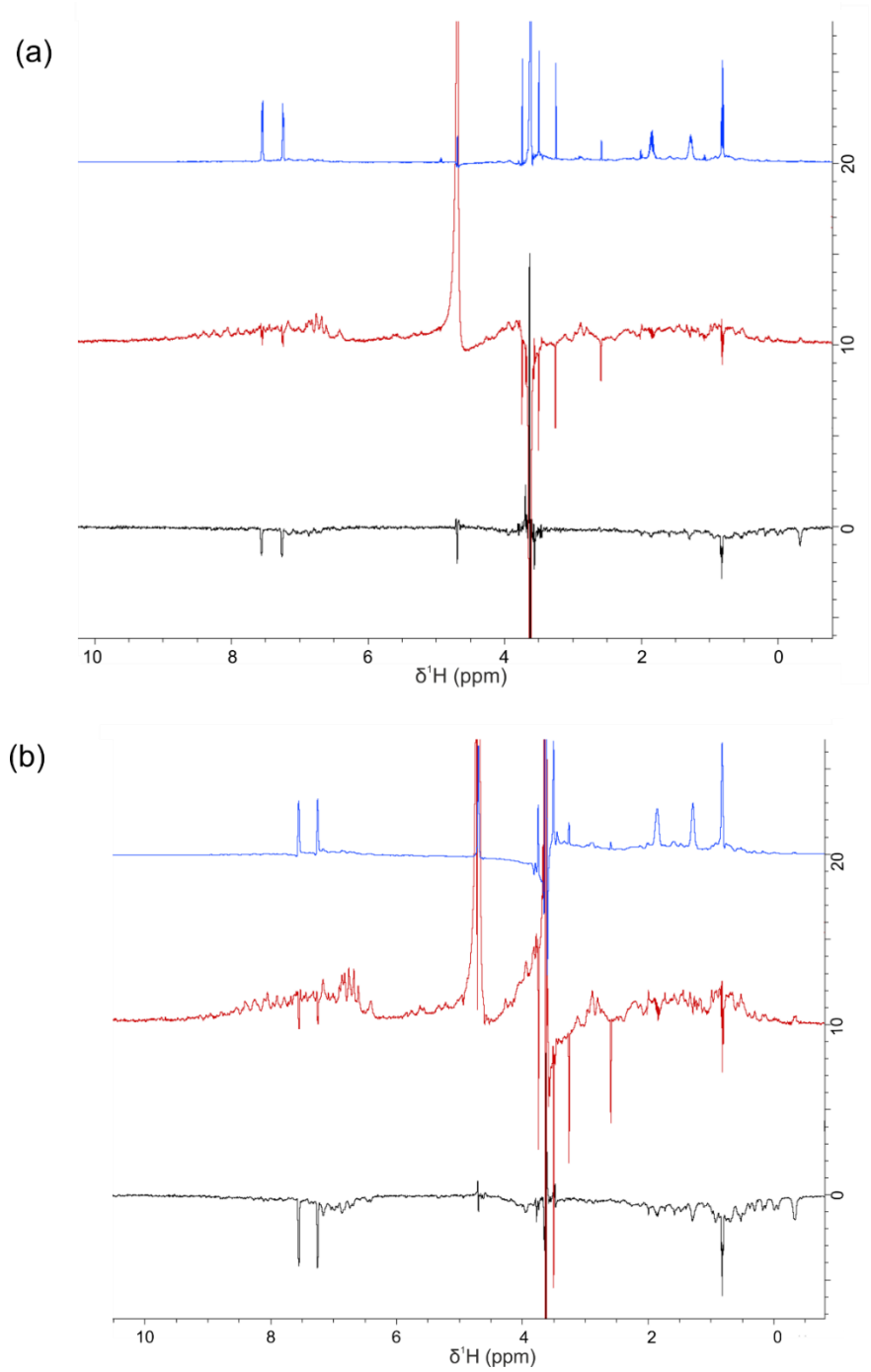


Figure A.6.7- 4.28, NCS1 and D2R peptide. (a) 1 mM **4.28** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand on NCS1 can be observed in the aromatic and aliphatic regions of the STD where the ligand signals appear in the STD and waterLOGSY where their corresponding signals in the waterLOGSY are reduced or positive. **(b)** 1 mM **4.28** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

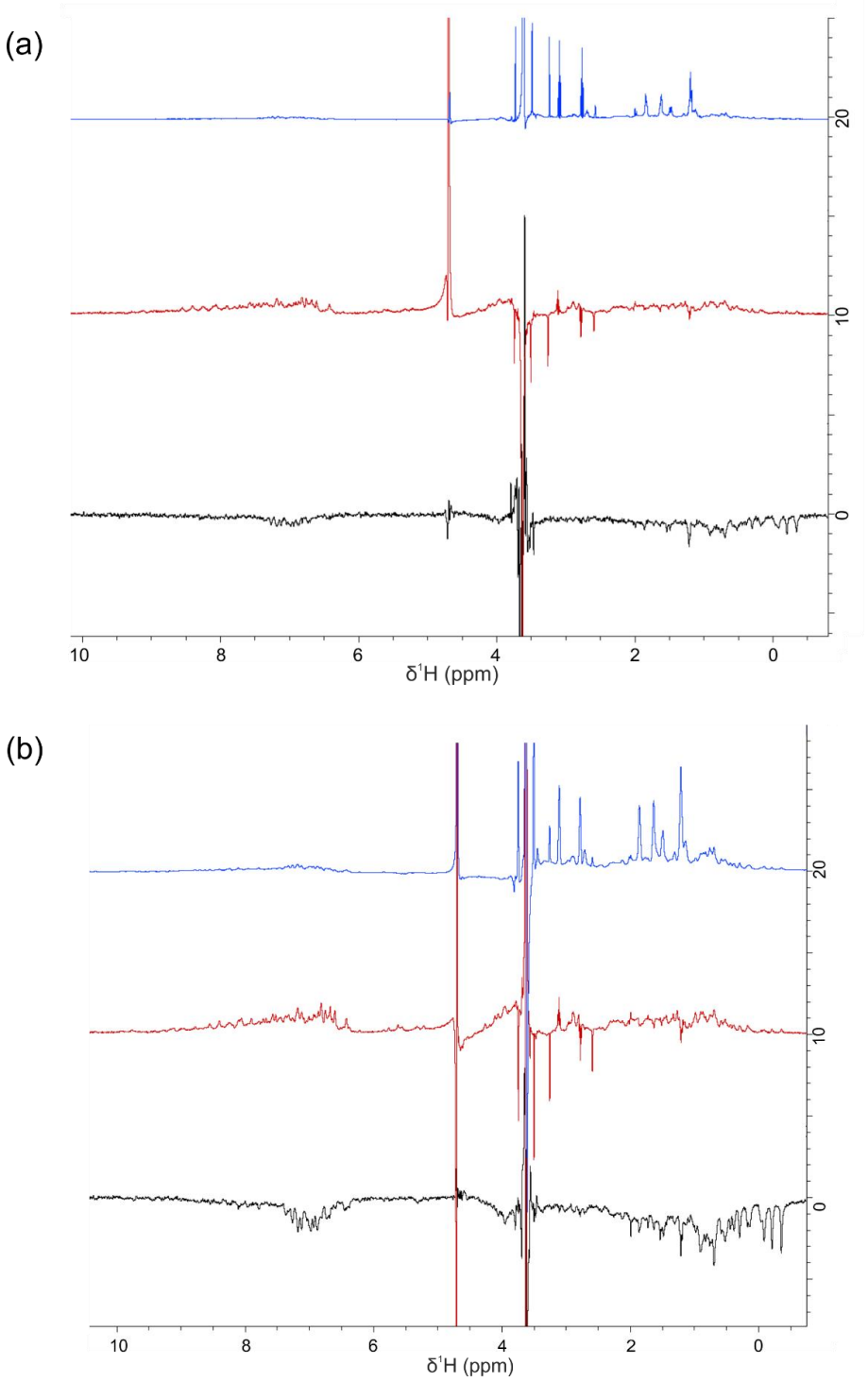


Figure A.6.8- 4.31, NCS1 and D2R peptide. (a) 1 mM **4.31** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. A possible weak effect was observed up-field in the aliphatic region of the STD. This was confirmed by reduced intensity of the corresponding signals in the waterLOGSY spectrum. **(b)** 1 mM **4.31** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

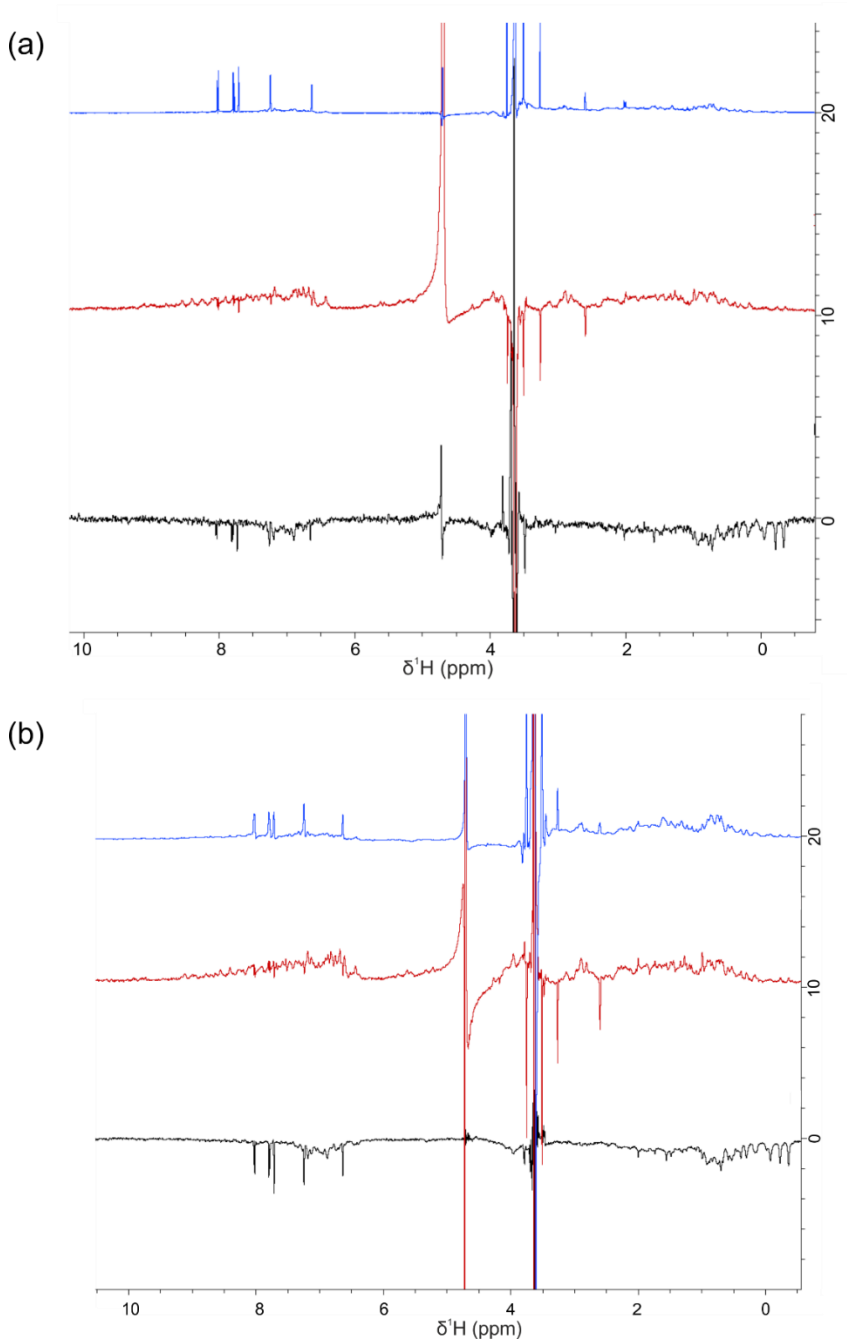


Figure A.6.9- 4.36, NCS1 and D2R peptide. (a) 1 mM **4.36** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand upon NCS1 can be observed in the aromatic region of the STD spectra where the ligand signals are apparent and waterLOGSY experiment where the corresponding signals are reduced or positive. **(b)** 1 mM **4.36** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.

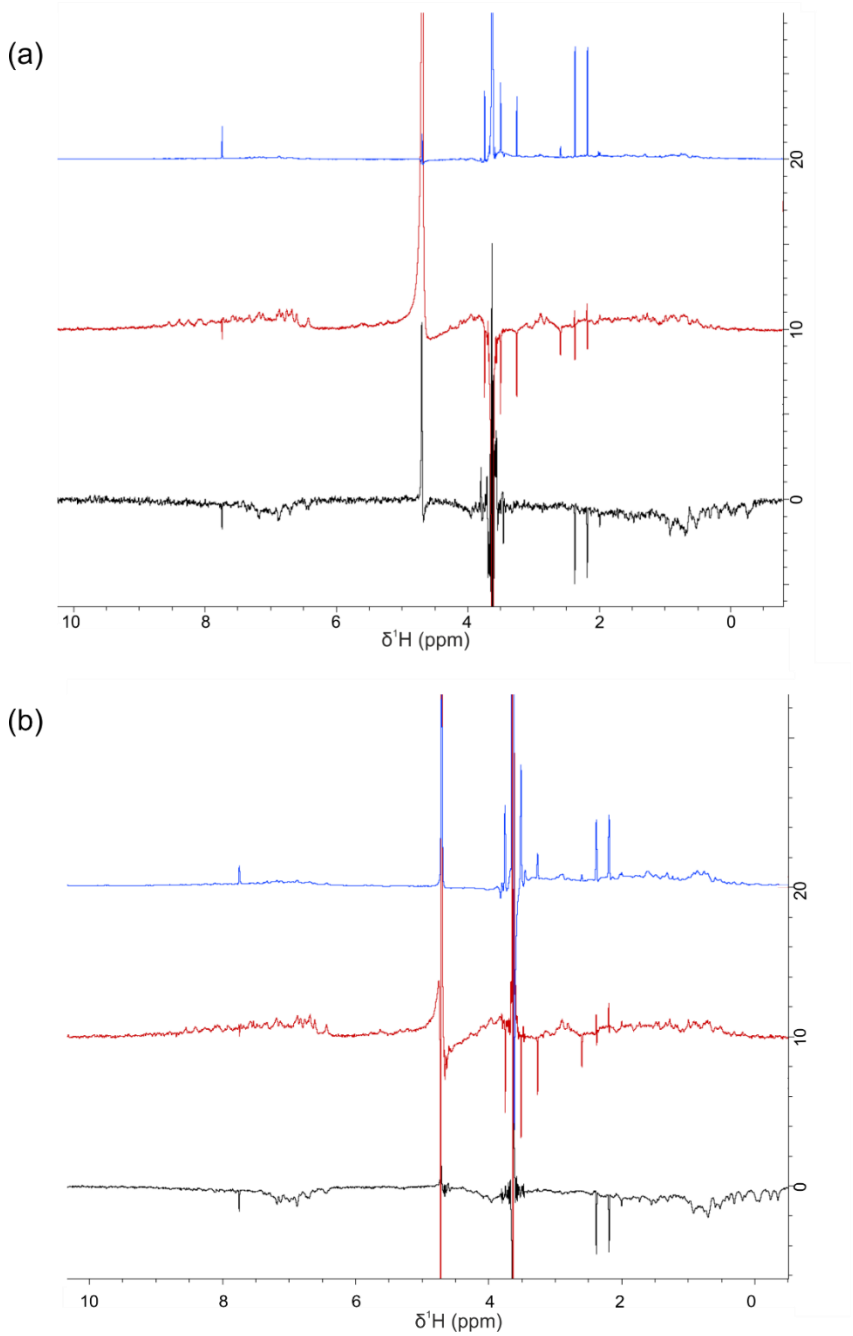


Figure A.6.10- 4.41, NCS1 and D2R peptide. (a) 1 mM **4.41** and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. An effect of the ligand upon NCS1 can be observed in the up-field aliphatic region of the STD spectra where some of the ligand signals are apparent and waterLOGSY experiment where the corresponding signals are reduced or positive. **(b)** 1 mM **4.41** with 50 μM D2R peptide and 50 μM NCS1, 1-D ^{1}H spectra seen in blue, waterLOGSY spectrum red and STD experiment black. No apparent change in the STD or waterLOGSY profiles when compared to the previously collected ligand and NCS1 spectra.